



**HAL**  
open science

# Two problems of statistical estimation for stochastic processes

Samvel Gasparyan

► **To cite this version:**

Samvel Gasparyan. Two problems of statistical estimation for stochastic processes. Statistics [math.ST]. Le Mans Université; Université d'État d'Erevan (Erevan, Arménie), 2016. English. NNT : 2016LEMA1031 . tel-01648364

**HAL Id: tel-01648364**

**<https://theses.hal.science/tel-01648364v1>**

Submitted on 25 Nov 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Thèse de Doctorat



# Samvel GASPARYAN

*Mémoire présenté en vue de l'obtention du  
grade de Docteur de l'Université du Maine  
sous le sceau de l'Université Bretagne Loire*

**École doctorale :** *Sciences et Technologies de l'Information, Mathématiques*

**Discipline :** *Mathématiques et leurs interactions*

**Spécialité :** *Statistiques*

**Unité de recherche :** *Laboratoire Manceau de Mathématiques*

**Soutenue le** *12 Décembre, 2016*

## Deux problèmes d'estimation statistique pour les processus stochastiques

### JURY

Rapporteurs : **Yury GOLUBEV**, Directeur de Recherche, CNRS  
**Masayuki UCHIDA**, Professeur, Osaka University, Japan

Examineurs : **Dominique DEHAY**, Professeur, Université Rennes 2  
**Saïd HAMADENE**, Professeur, Université du Maine  
**Marina KLEPTSYNA**, Professeur, Université du Maine  
**Sergueï PERGAMENHTCHIKOV**, Professeur, Université de Rouen

Directeur de Thèse : **Yury KUTOYANTS**, Professeur, Université du Maine  
Co-directeur de Thèse : **Victor OHANYAN**, Professeur, Université d'Etat d'Erevan, Arménie



# Contents

<b>Introduction</b>	<b>7</b>
<b>1 Estimation of the mean of a Poisson process</b>	<b>19</b>
1.1 Introduction . . . . .	19
1.2 Model description . . . . .	20
1.3 First Order Efficiency . . . . .	23
1.4 Second Order Efficiency . . . . .	30
1.4.1 Presentation of results . . . . .	30
1.4.2 Proof of Proposition 1.4 . . . . .	32
1.4.3 Proof of Proposition 1.3 . . . . .	37
<b>2 Approximation of the solution of a BSDE</b>	<b>51</b>
2.1 Stochastic differential equations . . . . .	51
2.2 Backward stochastic differential equations . . . . .	52
2.2.1 Markovian case . . . . .	52
2.2.2 Black and Scholes model . . . . .	54
2.3 Statement of the problem . . . . .	54
2.3.1 Continuous time observations . . . . .	54
2.3.2 High frequency asymptotics . . . . .	57
2.4 Parameter estimation . . . . .	58
2.4.1 Minimum contrast estimator . . . . .	58
2.4.2 Local asymptotic mixed normality . . . . .	60
2.5 Approximation of the solution . . . . .	62
2.5.1 Estimator process . . . . .	62
2.5.2 One-step maximum likelihood estimator (MLE) process . . . . .	63
2.5.3 Local asymptotic mixed normality . . . . .	64
2.5.4 Lower bound on approximation of the solution of a FBSDE . . . . .	67
2.5.5 Asymptotically efficient approximation of the solution of a FBSDE . . . . .	76
2.5.6 Pearson diffusion . . . . .	79

Contents	4
<b>Conclusion</b>	<b>85</b>
<b>Appendix</b>	<b>89</b>
<b>References</b>	<b>95</b>

## Remerciements

J'ai écrit cette thèse pendant longtemps, pourtant je n'ai pas écrit la thèse que je voulais. Je l'imaginai, cette thèse, autrement; j'avais beaucoup d'idées que je n'ai pas réussies à réaliser. Il y avait plusieurs instants où je voulais même abandonner ce projet, mais j'avais toujours des gens qui me soutenaient, m'encourageaient et me poussaient vers l'avant. Donc finalement j'ai réussi à la terminer et à vous de juger si c'était vraiment une réussite.

Si je n'ai pas abandonné cette thèse et si j'ai eu l'opportunité de la commencer et de la poursuivre c'est grâce à mes deux directeurs de thèse Youri A. Koutoyants et Victor K. Ohanyan, qui ont organisé mes études en France et que je veux remercier. Grâce à leur collaboration, à l'amitié qu'il y a entre eux j'ai eu cette grande opportunité de recherche et d'enseignement. Je veux remercier aussi la famille de Y.A. Koutoyants, surtout sa femme Nadejda, qui ont organisé pour moi une réception familiale en France et grâce à eux mon séjour en France, si loin de mon pays natal, n'était pas si difficile que l'on peut imaginer. Mes remerciements sincères aux rapporteurs qui ont bien lu et commenté mon travail. Je vous remercie et je vous assure que votre critique était très importante pour moi, j'en tiendrai compte. M. Youri Golubev avait vraiment une grande influence sur mon développement scientifique. Le premier chapitre de cette thèse est entièrement dédié à un problème initié par lui et son co-auteur, donc je passais pas mal de temps à lire ces travaux. Pour moi c'est un honneur qu'il ait accepté d'être un rapporteur pour ma thèse. J'avais l'opportunité de discuter de mon travail avec lui pendant la conférence S.A.P.S. X (Mars, 2015, Le Mans), ensuite j'ai profité d'une chance d'assister à une conférence organisée par M. Golubev à CIRM (février, 2016, Marseille). Je veux le remercier pour ces deux opportunités aussi. Je remercie mon deuxième rapporteur Masayuki Uchida, pour son accord d'être rapporteur et d'avoir fait une lecture attentive de ma thèse.

Pour mon développement scientifique je veux remercier aussi professeur Arnak Dalalyan. Mes remerciements particuliers à professeur Serguei Dachian pour sa bonne volonté de me toujours aider.

Je veux remercier les membres du jury Dominique Dehay, Saïd Hamadène, Marina Kleptsyna et Serguei Pergamenchtchikov qui ont accepté d'être examinateurs de ma soutenance. Depuis deux ans j'enseignais, donc je profitais de l'aide et du soutien de mes collègues. Leur contribution à mon développement comme chercheur-enseignant est bien appréciée. La liste commence par Marina Kleptsyna et Alex Popier, qui étaient mes «directeurs d'enseignement», et professeur Kleptsyna même m'a très catégoriquement forcé à l'époque à oublier l'idée d'abandonner cette thèse. La liste comprend Guillaume Voisin, Ghislain Dupont, Sylvain Maugeais, Youssef Serrestou (Analyse), Gérard Leloup (Algèbre Linéaire), Samir Ben Hariz (Statistiques Descriptives), encore une fois Alex Popier (Outils Mathématiques) et Alex Brouste (Statistique de Master 1) pour son aide constante de programmation en R. J'apprécie aussi mes collègues avec qui je discutais vraiment beaucoup, ce sont Wissal Sabbagh, avec qui je partageais mon bureau et Anastasiia Motrunich. Je mentionne ici les autres membres de notre laboratoire qui étaient vraiment de bonnes connaissances pour moi, ce sont Ivan Suarez, Christian Farinetta, Christophe Dutang, Marie-Amélie Morlais. Ma liste inclut aussi les professeurs de deux laboratoires qui me soutenaient toujours. En Arménie ce sont Gegham Gevorgyan et Artur Sahakyan, au Mans ce sont Laurent Denis, Saïd Hamadène et Anis Matoussi.

Bien sûr, je remercie mes amis doctorants (par ordre de soutenance)- Julie, Lin, Rui, Zhao, Chunhao, Achref, Lambert, Maroua, Alioune, Fanny, Arij, Rim et Dima. J'apprécie aussi une très bonne connaissance de Paris, Adélaïde Olivier. Enfin je dis un grand merci à Brigitte Bougard et Irène Croset pour leur aide au laboratoire. Je remercie aussi le professeur de français qui m'a aidé avec l'écriture. Je vous remercie Cécile.

Je remercie ma mère Ashkhen, mon père Bagrat, mon frère Manvel, toute ma grande famille et tous mes amis en Arménie.

Je remercie tous mes étudiants de leur gentille présence à mes cours.

Pour le financement de mes études en France je remercie le Gouvernement Français pour la bourse. J'ai profité aussi des aides financières de déplacement de deux universités que je représente; l'Université du Maine et l'Université d'État d'Erevan.

*20 mai, 2016, vendredi, Le Mans, France*

# Introduction

...all models are wrong, but some are useful.

---

**Box G. E. P.**

The first chapter is devoted to a problem of non-parametric estimation. The difference from parametric estimation is that the unknown object to be estimated is infinite dimensional. As a model we have continuous time observations of a periodic inhomogeneous Poisson process with the known period  $\tau$ . We suppose that the observation is done in the time interval  $[0, T]$ , where  $T = n\tau$ . The setting of the problem is asymptotic, that is, we will consider the case where  $n$  tends to infinity. A Poisson process is a point process - random countable realizations in some space, called the state space. We will identify the Poisson process with the counting process of the points of a realization in bounded subsets of the state space. Hence, the Poisson process is a non-negative, integer-valued process. To each Poisson process is associated a measure on the state space, called the mean measure of the Poisson process. If the state space is the interval  $[0, T]$ , then each locally-compact Borel measure can be given by a single function, which, for the mean measure we call the mean function. If the mean function  $\Lambda(\cdot)$  is absolutely continuous with respect to the Lebesgue measure

$$\Lambda(t) = \int_0^t \lambda(s) ds,$$

then the function  $\lambda(\cdot)$  is called the intensity function of the Poisson process. Periodicity of the Poisson process means that the intensity function exists and is a  $\tau$ -periodic function

$$\lambda(t) = \lambda(t + k\tau), k \in \mathcal{Z}_+, t \in [0, \tau].$$

The periodic Poisson process  $\{X(t), t \in [0, T], T = n\tau\}$  with the known period  $\tau$  can be transformed to a Poisson process on the interval  $[0, \tau]$ , observed indepen-



dently  $n$  times, using the following notations

$$\begin{aligned} X_j(t) &= X((j-1)\tau + t) - X((j-1)\tau), \quad t \in [0, \tau], \\ \mathbf{X}_j &= \{X_j(t), t \in [0, \tau]\}, \quad j = 1, \dots, n. \end{aligned}$$

Hence, our observations will be i.i.d. (independent, identically distributed) Poisson processes  $\mathbf{X}^n = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$  on the interval  $[0, \tau]$ . We consider the problem of the estimation of the mean function. The mean function estimation problem is very close to the distribution function estimation problem from i.i.d. observations of real random variables. More precisely, we can construct consistent estimators without regularity conditions on the unknown object (since there is an unbiased estimator for the unknown object). It is well known that, for example, in the density (i.i.d. random variables case) or intensity (inhomogeneous Poisson process) estimation problems, even for constructing a consistent estimator we have to impose regularity conditions (existence and Hölder continuity of some derivative, [40]) on the unknown object. In the works of Kutoyants [21],[20] it was shown that the empirical mean function

$$\hat{\Lambda}(t) = \frac{1}{n} \sum_{j=1}^n X_j(t)$$

is consistent, asymptotically normal estimator with the optimal rate and is even asymptotically efficient for a large number of loss functions (including polynomials). Our work treats only the case of  $L_2$  loss function or known as mean integrated squared error (MISE). Asymptotic efficiency we understand in the following way. For all estimators the following lower bound is true

$$\liminf_{\delta \rightarrow 0} \liminf_{n \rightarrow +\infty} \sup_{\Lambda \in V_\delta} n \int_0^\tau \mathbf{E}_\Lambda (\bar{\Lambda}_n(s) - \Lambda(s))^2 ds \geq \int_0^\tau \Lambda^*(s) ds,$$

where  $V_\delta$  is a shrinking vicinity of  $\Lambda^*$  in sup norm

$$V_\delta = \{\Lambda(\cdot) : \sup_{t \in [0, \tau]} |\Lambda(t) - \Lambda^*(t)| \leq \delta\}.$$

The estimator  $\tilde{\Lambda}_n(t)$  for which we have equality

$$\liminf_{\delta \rightarrow 0} \liminf_{n \rightarrow +\infty} \sup_{\Lambda \in V_\delta} n \int_0^\tau \mathbf{E}_\Lambda (\tilde{\Lambda}_n(s) - \Lambda(s))^2 ds = \int_0^\tau \Lambda^*(s) ds$$

is said to be asymptotically efficient.

We will use this result in a slightly modified way

$$\inf_{\Lambda_n} \liminf_{n \rightarrow +\infty} \sup_{\Lambda \in \mathcal{F}} \left( \mathbf{E}_\Lambda \|\sqrt{n}(\bar{\Lambda}_n - \Lambda)\|^2 - \int_0^\tau \Lambda(t) dt \right) = 0,$$

where  $\mathcal{F}$  is a subset of a ball of radius  $R$  centered around a given (known) mean function  $\Lambda^*$ . The statement of asymptotic efficiency of the mean function is true for several bounded sets (particularly for  $\Sigma(R)$  and  $\mathcal{F}_m^{per}(R, S)$ , see below for definitions), and, obviously, if it is true for some set, then it is true for all sets containing that set. The empirical mean function is not the only asymptotically efficient estimator. Our first goal is to construct a class of asymptotically efficient estimators. Below we always consider a kernel type estimator

$$\tilde{\Lambda}_n(t) = \int_0^\tau K_n(s-t)(\hat{\Lambda}_n(s) - \Lambda_*(s))ds + \Lambda_*(t),$$

with non-negative, integrable, normalized kernels  $K_n(\cdot)$  on  $[-\frac{\tau}{2}, \frac{\tau}{2}]$ ,

$$K_n(u) \geq 0, \quad \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} K_n(u)du = 1, \quad n \in \mathcal{N},$$

that are symmetric and we continue them  $\tau$  periodically on the whole real line  $\mathbf{R}$

$$K_n(u) = K_n(-u), \quad K_n(u) = K_n(u + k\tau), \quad u \in \left[-\frac{\tau}{2}, \frac{\tau}{2}\right], \quad k \in \mathcal{Z}.$$

Then, their Fourier coefficients with respect to the trigonometric basis will be

$$\begin{aligned} \tilde{\Lambda}_{1,n} &= \hat{\Lambda}_{1,n}, \quad \tilde{\Lambda}_{2l,n} = \sqrt{\frac{\tau}{2}} K_{2l,n}(\hat{\Lambda}_{2l,n} - \Lambda_{2l}^*) + \Lambda_{2l}^*, \\ \tilde{\Lambda}_{2l+1,n} &= \sqrt{\frac{\tau}{2}} K_{2l,n}(\hat{\Lambda}_{2l+1,n} - \Lambda_{2l+1}^*) + \Lambda_{2l+1}^*, \quad l \in \mathcal{N}, \end{aligned}$$

with

$$\Lambda_l = \int_0^\tau \Lambda(t)\phi_l(t)dt, \quad \Lambda_l^* = \int_0^\tau \Lambda_*(t)\phi_l(t)dt.$$

Our first result states

**Proposition** ([6]). *Suppose that the kernels satisfy also the condition*

$$n \sup_{l \geq 1} \left| \sqrt{\frac{\tau}{2}} K_{2l,n} - 1 \right|^2 \longrightarrow 0,$$

as  $n \rightarrow +\infty$ . Then, over a  $\Lambda_*$  centered  $\mathcal{B}(R) = \{\Lambda : \int_0^\tau [\Lambda(t) - \Lambda_*(t)]^2 dt \leq R\}$  ball of mean functions in  $L_2$ , the following equality holds

$$\lim_{n \rightarrow +\infty} \sup_{\Lambda \in \mathcal{B}(R)} \left( \mathbf{E}_\Lambda \|\sqrt{n}(\tilde{\Lambda}_n - \Lambda)\|^2 - \int_0^\tau \Lambda(t)dt \right) = 0.$$

We prove also

**Proposition.** *For the kernel-type estimator with kernels satisfying the condition*

$$n \sup_{l \geq 1} \left| \frac{\sqrt{\frac{\tau}{2}} K_{2l,n} - 1}{\frac{2\pi l}{\tau}} \right|^2 \longrightarrow 0,$$

as  $n \rightarrow +\infty$ , the following equality holds

$$\lim_{n \rightarrow +\infty} \sup_{\Lambda \in \Sigma(R)} \left( \mathbf{E}_\Lambda \|\sqrt{n}(\tilde{\Lambda}_n - \Lambda)\|^2 - \int_0^\tau \Lambda(t) dt \right) = 0,$$

where the set  $\Sigma(R)$  is

$$\Sigma(R) = \left\{ \Lambda, \int_0^\tau (\lambda(t) - \lambda_*(t))^2 dt \leq R, \Lambda_*(0) = \Lambda(0) = 0, \Lambda_*(\tau) = \Lambda(\tau) \right\}.$$

To give explicitly a class of asymptotically efficient estimators

**Remark.** *Suppose that  $K(u)$ ,  $u \in [-\frac{\tau}{2}, \frac{\tau}{2}]$  is a non-negative, integrable, normalized function*

$$K(u) \geq 0, u \in \left[-\frac{\tau}{2}, \frac{\tau}{2}\right], \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} K(u) du = 1,$$

which is symmetric and we continue it  $\tau$  periodically on the whole real line

$$K(u) = K(-u), \quad K(u) = K(u + k\tau), \quad u \in \left[-\frac{\tau}{2}, \frac{\tau}{2}\right], \quad k \in \mathcal{Z}.$$

Let the positive sequence  $h_n \leq 1$  be such that  $h_n^2 n \rightarrow 0$ ,  $n \rightarrow +\infty$ . Then, the kernels

$$K_n(u) = \frac{1}{h_n} K\left(\frac{u}{h_n}\right) \mathbb{I}\left\{|u| \leq \frac{\tau}{2} h_n\right\}$$

satisfy the condition

$$n \sup_{l \geq 1} \left| \frac{\sqrt{\frac{\tau}{2}} K_{2l,n} - 1}{\frac{2\pi l}{\tau}} \right|^2 \longrightarrow 0.$$

To compare asymptotically efficient estimators we prove the following result. Let

$$\mathcal{F}_m^{per}(R, S) = \left\{ \Lambda(\cdot) : \int_0^\tau [\Lambda^{(m)}(t)]^2 dt \leq R, \Lambda(0) = 0, \Lambda(\tau) = S \right\},$$

where  $R > 0$ ,  $S > 0$ ,  $m > 1$ ,  $m \in \mathcal{N}$  are given constants. Periodicity means that the intensity function  $\lambda(\cdot)$  is periodic (hence the equality of its values and the values of its derivatives on the endpoints of the interval  $[0, \tau]$ ). Introduce as well

$$\Pi = \Pi_m(R, S) = (2m - 1)R \left( \frac{S}{\pi R (2m - 1)(m - 1)} \right)^{\frac{2m}{2m-1}}.$$

**Proposition.** *Then, for all estimators  $\bar{\Lambda}_n(t)$  of the mean function  $\Lambda(t)$ , following lower bound holds*

$$\lim_{n \rightarrow +\infty} \sup_{\Lambda \in \mathcal{F}_m^{per}(R,S)} n^{\frac{2m}{2m-1}} \left( \int_0^\tau \mathbf{E}_\Lambda (\bar{\Lambda}_n(t) - \Lambda(t))^2 dt - \frac{1}{n} \int_0^\tau \Lambda(t) dt \right) \geq -\Pi.$$

This inequality compares second order asymptotic term of maximal loss over some non-parametric class of functions (under additional regularity conditions on the unknown mean function). Hence the estimators reaching that lower bound will be called second order efficient. In our work we explicitly calculate asymptotic minimal error for the second order estimation. The constant  $\Pi_m(R, S)$  plays the same role in second order estimation as the Pinsker constant in density estimation problem or the inverse of the Fisher information in the regular parametric estimation problems. But unlike mentioned problems here the constant is negative. This is due to the fact that for the empirical mean function

$$\mathbf{E}_\Lambda \int_0^\tau (\hat{\Lambda}(t) - \Lambda(t))^2 dt = \frac{1}{n} \int_0^\tau \Lambda(t) dt,$$

therefore the second term is equal to zero. We propose also an estimator

$$\Lambda_n^*(t) = \hat{\Lambda}_{0,n} \phi_0(t) + \sum_{l=1}^{+\infty} K_{l,n} \hat{\Lambda}_{l,n} \phi_l(t),$$

where  $\{\phi_l\}_{l=1}^{+\infty}$  is the trigonometric cosine basis on  $L_2[0, \tau]$ . For the first order efficiency we used the trigonometric basis with both sine and cosine functions present, here we took only the cosine basis to avoid technical difficulties and, as a result, the proposed estimator does not depend on the center of the ellipsoid (here it is present as the quantity  $S$ ).  $\hat{\Lambda}_{l,n}$  are the Fourier coefficients of the empirical mean function with respect to this basis and

$$\tilde{K}_{l,n} = \left( 1 - \left| \frac{\pi l}{\tau} \right|^m \alpha_n^* \right)_+, \quad \alpha_n^* = \left[ \frac{S \tau}{nR \pi} \frac{m}{(2m-1)(m-1)} \right]^{\frac{m}{2m-1}},$$

$$N_n = \frac{\tau}{\pi} (\alpha_n^*)^{-\frac{1}{m}} \approx C n^{\frac{1}{2m-1}}, \quad x_+ = \max(x, 0), \quad x \in \mathbf{R}.$$

We emphasize the fact that though in the definition there is an infinite sum, but in reality it contains only finite terms, but the number of terms  $N_n$  tends to infinity as  $n \rightarrow +\infty$  with the rate  $n^{\frac{1}{2m-1}}$ . Our next result is

**Proposition.** *The estimator  $\Lambda_n^*(t)$  attains the lower bound described above, that is,*

$$\lim_{n \rightarrow +\infty} \sup_{\Lambda \in \mathcal{F}_m^{per}(R,S)} n^{\frac{2m}{2m-1}} \left( \int_0^\tau \mathbf{E}_\Lambda (\Lambda_n^*(t) - \Lambda(t))^2 dt - \frac{1}{n} \int_0^\tau \Lambda(t) dt \right) = -\Pi.$$

That is, our proposed estimator is asymptotically second order efficient. Result is presented in [6].

The explicit calculation of the asymptotic error in a non-parametric estimation problem was first done by Pinsker [36] in the model of observation of a signal in the white Gaussian noise. The idea was to consider the minimax risk of integral-type quadratic loss functions on a Sobolev ellipsoid. The concept of second order efficiency was introduced by Golubev and Levit [13] in the problem of distribution function estimation for the i.i.d. model. In the paper [13], authors proved a lower bound which allows to compare second term of the expansion of the maximal loss over some set of functions and minimize that term. They proposed also an estimator which attains that lower bound, hence that lower bound is sharp.

Later, second order efficiency was considered for some other models. For example, Dalalyan and Kutoyants [3] proved second order asymptotic efficiency in the estimation problem of the invariant density of an ergodic diffusion process. Golubev and Härdle [12] proved second order asymptotic efficiency in partial linear models.

The second chapter of this work is devoted to the approximation problem of the solution of a forward-backward stochastic differential equation (forward BSDE or FBSDE). We suppose that the diffusion coefficient of the forward equation depends on an unknown one-dimensional parameter, therefore the solution of the backward equation also depends on that parameter. It is well-known that if we observe the solution of a stochastic differential equation on the whole interval, even bounded and very small, then we can estimate the unknown parameter in the diffusion coefficient “without an error”, hence it is not a statistical problem. Our considerations are based on the observations of the solution of the forward equation on a finite interval, at equidistant discrete points. As the number of observations tends to infinity, the distance between the observation times tends to zero. Such a statement of problem is called *high frequency asymptotics*. At a given point of time we have to construct an estimator for the unknown parameter based on the observation times before that time only, hence we have to construct an estimator process. We are seeking an estimator which is computationally simple, but of course, we do not want to lose in the performance of the estimator (we want it to be still asymptotically efficient). For that reason we are looking for an estimator which is also asymptotically efficient.

The BSDE was first introduced in the linear case by Bismuth [2] and in general case this equation was studied by Pardoux and Peng [33]. Since that time the BSDE attracts attention of probabilists working in financial mathematics and has obtained an intensive development (see, e.g. El Karoui *et al.* [5], Ma and Yong [30] and the references therein). The detailed exposition of the current state of

this theory can be found in Pardoux and Răscanu [35].

We consider the following problem. Suppose that we have a stochastic differential equation (called *forward*)

$$dX_t = S(t, X_t) dt + \sigma(\vartheta, t, X_t) dW_t, \quad X_0, \quad 0 \leq t \leq T,$$

and two functions  $f(t, x, y, z)$  and  $\Phi(x)$  are given. We have to find a couple of stochastic processes  $(Y_t, Z_t)$  such that it satisfies the stochastic differential equation (called *backward*)

$$dY_t = -f(t, X_t, Y_t, Z_t) dt + Z_t dW_t, \quad 0 \leq t \leq T$$

with the final value  $Y_T = \Phi(X_T)$ .

The solution of this problem is well-known. We have to solve a special partial differential equation, to find its solution  $u(t, x, \vartheta)$  and to put  $Y_t = u(t, X_t, \vartheta)$  and  $Z_t = \sigma(\vartheta, t, X_t) u'_x(t, X_t, \vartheta)$ .

We are interested in the problem of approximation of the solution  $(Y_t, Z_t)$  in the situation where the parameter  $\vartheta$  is unknown. Therefore we first estimate this parameter with help of some good estimator  $\vartheta_{t,n}^*$ ,  $0 < t \leq T$  based on the discrete time observations (till time  $t$ ) of the solution of the forward equation and then we propose the approximations  $Y_t^* = u(t, X_t, \vartheta_{t,n}^*)$ ,  $Z_t^* = \sigma(\vartheta_{t,n}^*, t, X_t) u'_x(t, X_t, \vartheta_{t,n}^*)$ . Moreover, we show that the proposed approximations are in some sense asymptotically optimal.

The main difficulty in the construction of this approximation is to find an estimator-process  $\vartheta_{t,n}^*$ ,  $0 < t \leq T$  which can be easily calculated for all  $t \in (0, T]$  and at the same time has asymptotically optimal properties. Unfortunately we cannot use the well-studied pseudo-MLE (maximum likelihood estimator) based on the pseudo-maximum likelihood function because its calculation is related to the solution of nonlinear equations and numerically is sufficiently difficult problem.

The one-step MLE was proposed in the local asymptotically normal statistical models by Le Cam [25] as a method to improve an arbitrary estimator with the optimal rate up to an asymptotically efficient one. Volatility parameter estimation has another asymptotic property. Under regularity conditions, the volatility parameter estimation model is locally asymptotically mixed normal ([4], [9], [11]). We propose here a one-step MLE-process, which was recently introduced in the case of ergodic diffusion [23] and diffusion process with small noise [24]. As in the construction of the one-step MLE, we take a preliminary estimator and improve its asymptotic performance by transforming it to an optimal estimator. The difference is that the one-step MLE-process allows us even to improve the rate

of the preliminary estimator. Note that the multi-step Newton-Raphson procedure for construction of parameter estimators was first proposed by Kamatani and Uchida [19]. They considered the discrete time observations in the case of observations of ergodic diffusion process where the unknown parameters were in the drift and diffusion coefficients. The asymptotics in their work corresponds to the “high frequency” and large samples. The multi-grid process allows them to obtain asymptotically efficient estimators. The review of statistical problems for the BSDE model of observations can be found in [22].

Note that the problem of volatility parameter estimation by discrete time observations is actually a well developed branch of statistics (see, for example, [38] and references therein). The particularity of our approach is due to the need of updated *on-line* estimator  $\vartheta_{t,n}^*$  which depends on the first observations up to time  $t$ .

Let us fix some (small)  $\tau > 0$ . We call the interval  $[0, \tau]$  *the learning interval* and we construct the estimator process for the values of  $t \in [\tau, T]$ . Based on the learning interval we construct a preliminary estimator, then we improve this estimator up to an optimal (asymptotically efficient) estimator which, on the other hand, is computationally easy calculable. As a preliminary estimator  $\hat{\vartheta}_{\tau,n}$  we take a particular *minimum contrast estimator* (MCE) (for a general method of constructing MCE estimators for the diffusion parameter see [9]) which is called the pseudo-maximum likelihood estimator (PMLE), constructed by the observations  $X^{\tau,n} = (X_0, X_{t_{1,n}}, \dots, X_{t_{N,n}})$ , where  $t_{N,n} \leq \tau < t_{N+1,n}$ . To defining the PMLE introduce the log pseudo-likelihood ratio

$$L_{t,k}(\vartheta, X^k) = -\frac{1}{2} \sum_{j=0}^k \ln [2\pi\sigma^2(\vartheta, t_{j-1}, X_{t_{j-1}}) \delta] - \sum_{j=1}^k \frac{[X_{t_j} - X_{t_{j-1}} - S(t_{j-1}, X_{t_{j-1}}) \delta]^2}{2\sigma^2(\vartheta, t_{j-1}, X_{t_{j-1}}) \delta}$$

and define the PMLE  $\hat{\vartheta}_{\tau,n}$  by the equation

$$L_{\tau,N}(\hat{\vartheta}_{\tau,n}, X^N) = \sup_{\vartheta \in \Theta} L_{\tau,N}(\vartheta, X^N), \quad .$$

This estimator is consistent and asymptotically conditionally normal ([9])

$$\begin{aligned} \sqrt{\frac{n}{T}} \left( \hat{\vartheta}_{\tau,n} - \vartheta_0 \right) &= \mathbb{I}_{\tau,n}(\vartheta_0)^{-1} \sqrt{2} \sum_{j=1}^N \frac{\dot{\sigma}(\vartheta_0, t_{j-1}, X_{t_{j-1}})}{\sigma(\vartheta_0, t_{j-1}, X_{t_{j-1}})} w_j + o(1) \\ &\implies \xi_{\tau}(\vartheta_0) = \mathbb{I}_{\tau}(\vartheta_0)^{-1} \sqrt{2} \int_0^{\tau} \frac{\dot{\sigma}(\vartheta_0, s, X_s)}{\sigma(\vartheta_0, s, X_s)} dw(s). \end{aligned}$$

Here the random Fisher information matrix is

$$\mathbb{I}_\tau(\vartheta_0) = 2 \int_0^\tau \frac{\dot{\sigma}(\vartheta_0, s, X_s) \dot{\sigma}(\vartheta_0, s, X_s)^\top}{\sigma^2(\vartheta_0, s, X_s)} ds$$

and

$$\mathbb{I}_{\tau,n}(\vartheta_0) = 2 \sum_{j=1}^N \frac{\dot{\sigma}(\vartheta_0, t_{j-1}, X_{t_{j-1}}) \dot{\sigma}(\vartheta_0, t_{j-1}, X_{t_{j-1}})^\top}{\sigma^2(\vartheta_0, t_{j-1}, X_{t_{j-1}})} \delta \longrightarrow \mathbb{I}_\tau(\vartheta_0),$$

where *dot* means the derivative with respect to the unknown parameter  $\vartheta$ .

Introduce the pseudo score-function ( $A_{j-1}(\vartheta) = \sigma^2(\vartheta, t_{j-1}, X_{t_{j-1}})$ )

$$\begin{aligned} \Delta_{k,n}(\vartheta, X^k) &= \sum_{j=1}^k \dot{\ell}(\vartheta, X_{t_{j-1}}, X_{t_j}) \\ &= \sum_{j=1}^k \frac{\left[ (X_{t_j} - X_{t_{j-1}} - S_{j-1} \delta)^2 - A_{j-1}(\vartheta) \delta \right] \dot{A}_{j-1}(\vartheta)}{2A_{j-1}^2(\vartheta) \sqrt{\delta}}. \end{aligned}$$

For any  $t \in [\tau, T]$  define  $k$  by the condition  $t_k \leq t < t_{k+1}$ . The one-step MLE-process is introduced by the relation

$$\vartheta_{k,n}^* = \hat{\vartheta}_{\tau,n} + \sqrt{\delta} \mathbb{I}_{k,n}(\hat{\vartheta}_{\tau,n})^{-1} \Delta_{k,n}(\hat{\vartheta}_{\tau,n}, X^k), \quad k = N+1, \dots, n. \quad (0.1)$$

Our goal is to show that the corresponding approximation

$$Y_{t_k,n}^* = u(t_k, X_{t_k}, \vartheta_{k,n}^*), \quad k = N+1, \dots, n,$$

is asymptotically efficient. To do this we need to present the lower bound on the risks of all estimators and then to show that for the proposed approximation this lower bound is reached. Our first result is

**Proposition.** *The one-step MLE-process  $\vartheta_{k,n}^*, k = N+1, \dots, n$  is consistent, asymptotically conditionally normal (stable convergence)*

$$\delta^{-1/2} (\vartheta_{k,n}^* - \vartheta_0) \Longrightarrow \xi_t(\vartheta_0), \quad \xi_t(\vartheta_0) = \frac{\Delta_t(\vartheta_0)}{\mathbb{I}_t(\vartheta_0)}$$

and is asymptotically efficient for  $t \in [\tau_*, T]$  where  $\tau < \tau_* < T$  and for bounded loss functions.

Then we prove a lower bound for the approximation of the solution of a FBSDE



**Proposition.** *Suppose that the coefficients of the diffusion process satisfies  $\mathcal{R}$  conditions, then, for the loss function  $\ell(u) = |u|^p$ ,  $p > 0$ , the following lower bound is true*

$$\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow +\infty} \sup_{|\vartheta - \vartheta_0| < \varepsilon} \mathbf{E}_{\vartheta} \ell(\delta^{-1/2} (\bar{Y}_{t_k, n} - Y_{t_k})) \geq \mathbf{E}_{\vartheta_0} \ell(\dot{u}(\vartheta_0, t, X_t) \xi_t(\vartheta_0)).$$

Here  $u(\vartheta, t, x)$  satisfies the equation

$$\frac{\partial u}{\partial t} + S(t, x) \frac{\partial u}{\partial x} + \frac{\sigma(\vartheta, t, x)^2}{2} \frac{\partial^2 u}{\partial x^2} = -f\left(t, x, u, \sigma(\vartheta, t, x) \frac{\partial u}{\partial x}\right).$$

**Proposition.** *Suppose that the conditions of regularity hold, then the estimators*

$$Y_{t_k, n}^* = u(t_k, X_{t_k}, \vartheta_{k, n}^*), \quad Z_{t_k, n}^* = u'_x(t_k, X_{t_k}, \vartheta_{k, n}^*) \sigma(t_k, X_{t_k}, \vartheta_{k, n}^*), \quad t_k \in [\tau, T],$$

are consistent  $Y_{t_k, n}^* \rightarrow Y_{t_k}$ ,  $Z_{t_k, n}^* \rightarrow Z_{t_k}$ , (convergence in probability) and asymptotically conditionally normal (stable convergence)

$$\begin{aligned} \delta^{-1/2} (Y_{t_k, n}^* - Y_{t_k}) &\Longrightarrow \langle \dot{u}(t, X_t, \vartheta_0), \xi_t(\vartheta_0) \rangle, \\ \delta^{-1/2} (Z_{t_k, n}^* - Z_{t_k}) &\Longrightarrow \sigma(t, X_t, \vartheta_0) \langle \dot{u}'_x(t, X_t, \vartheta_0), \xi_t(\vartheta_0) \rangle \\ &\quad + u'_x(t, X_t, \vartheta_0) \langle \dot{\sigma}(t, X_t, \vartheta_0), \xi_t(\vartheta_0) \rangle. \end{aligned}$$

These results are presented in the work [7].

It have to be mentioned that we could construct the approximation of  $Y_t$  and  $Z_t$  as follows

$$\hat{Y}_{t, n} = u(t, X_t, \hat{\vartheta}_{k, N}) \text{ and } \hat{Z}_t = \sigma(\hat{\vartheta}_{k, N}, t, X_t) u'_x(t, X_t, \hat{\vartheta}_{k, N}),$$

that is, using only the preliminary estimator. Note that this type of approximation is not asymptotically efficient, since we use only part of the observations (only the learning interval). Hence we are looking for another estimator of  $\vartheta$  which can provide smaller error of estimation.

Then, we are considering a Pearson diffusion

$$dX_t = -X_t dt + \sqrt{\vartheta + X_t^2} dW_t, \quad X_0, \quad 0 \leq t \leq T.$$

For this model, using the preliminary estimator

$$\bar{\vartheta}_N = \frac{n}{TN} \left[ X_{t_N}^2 - X_0^2 - 2 \sum_{j=1}^N X_{t_{j-1}} [X_{t_j} - X_{t_{j-1}}] - \sum_{j=1}^N X_{t_{j-1}}^2 \delta \right],$$

we propose the one-step MLE process

$$\vartheta_{t_k,n}^* = \bar{\vartheta}_N + \sqrt{\delta} \sum_{j=1}^k \frac{[X_{t_j} - X_{t_{j-1}} + X_{t_{j-1}}\delta]^2 - (\bar{\vartheta}_N + X_{t_{j-1}}^2)\delta}{2\mathbb{I}_{t_k,n}(\bar{\vartheta}_N) (\bar{\vartheta}_N + X_{t_{j-1}}^2)^2 \sqrt{\delta}}, \quad \tau \leq t_k \leq T.$$

and prove the proposition

**Proposition.** *The one-step MLE-process  $\vartheta_{t_k,n}^*$  is consistent: for any  $\nu > 0$*

$$\mathbf{P}_{\vartheta_0} \left( \max_{N \leq k \leq n} |\vartheta_{t_k,n}^* - \vartheta_0| > \nu \right) \rightarrow 0$$

and for all  $t \in (\tau, T]$  the convergence

$$\delta^{-1/2} (\vartheta_{t_k,n}^* - \vartheta_0) \implies \zeta_t(\vartheta_0)$$

holds. Moreover, this estimator is asymptotically efficient.

The result is presented in [8].



# Chapter 1

## Estimation of the mean of a Poisson process

### 1.1 Introduction

As an introduction to the theory of Poisson processes can be used the book [18], for the introduction to the statistical theory of Poisson processes we refer to [21].

This chapter deals with the second order estimation problem for the mean function of a Poisson process. Poisson process is a special type of point processes which serves for modeling the occurrence of random events during a time interval, with the property that the number of events are stochastically independent on non-intersecting time intervals and have the Poisson distribution on these intervals. Being the simplest example of a point process with useful applications in modeling various random events (as for example, arrivals of clients in some serving locations) it serves as one of the building blocks for creating statistical concepts for stochastic processes. In this chapter, observing an inhomogeneous Poisson process on a segment of the real line independently several times, our goal is to estimate the mean function of that process in an asymptotically efficient way, as the number of observations tends to infinity. The problem has very evident resemblance to the estimation problem of the distribution function of a real random variable from independent, identically distributed observations. As in the latter problem, where empirical distribution function turns out to be an asymptotically efficient estimator in the minimax sense (see, for example, [10],[28]), the empirical mean function is an asymptotically the best estimator for the mean function for various loss functions ([21],[20]). To compare the performances of estimators throughout this chapter we use only the Hilbert space norm of square integrable functions and

are interested in minimizing the mean integrated square error (MISE) asymptotically over bounded sets. Obviously enough, the empirical mean function is not the only asymptotically efficient estimator for the mean function in this sense. We start by construction a class of estimators that share the same property over some specified compact set. The main result of this chapter consists of proving a lower bound, which compares the second order mean integrated square error (SMISE) of all estimators and allows to compare the asymptotically efficient (first order) estimators. The proof is done using the technique developed in [1], which uses the van Trees inequality from [10]. Thus, following the work [13], we introduce the notion of the second order asymptotic efficiency. This is done under additional smoothness assumptions on the unknown mean function. The rate of convergence of the second order term is explored and the limit constant of the SMISE is calculated. The latter plays the role of the Pinsker constant of nonparametric estimation problems where the rate of convergence is less than the classical parametric rate of convergence. The whole notion of the second order efficiency (introduced in [13] for the distribution function estimation and carried out for various other models in [3],[12]) serves as an analogue of the Pinsker theory [36] (see also [31],[40]) for the problems where the classical parametric rate of convergence is attained in a nonparametric estimation problem. In the last part of this chapter, the construction of the second order asymptotically efficient estimator is given.

## 1.2 Model description

We consider the mean function estimation problem of a periodic Poisson process with the known period  $\tau$ . Recall that  $\{X(t), t \in [0, T]\}$  is an inhomogeneous Poisson process if for each fixed time point it is a non-negative, integer-valued random variable, starts from zero  $X(0) = 0$  almost surely, has independent increments and there is a non-decreasing, continuous function  $\Lambda(t)$  such that

$$\mathbf{P}(X(t) - X(s) = k) = \frac{[\Lambda(t) - \Lambda(s)]^k}{k!} e^{-[\Lambda(t) - \Lambda(s)]}, \quad 0 \leq s < t \leq T, k \in \mathcal{Z}_+,$$

( $\mathcal{Z}_+$  denotes the set of all non-negative integer numbers). Here  $\Lambda(t)$ ,  $t \in [0, \tau]$  is called mean function of the Poisson process, because  $\mathbf{E}X(t) = \Lambda(t)$ . For each non-decreasing, continuous function there is a Poisson process with the mean function being that function ([18]), hence the mean functions will be identified with these functions. If the mean function is absolutely continuous

$$\Lambda(t) = \int_0^t \lambda(s) ds,$$

then  $\lambda(\cdot)$  is called the intensity function. Periodicity of the Poisson process means that the intensity function exists and is a  $\tau$ -periodic function

$$\lambda(t) = \lambda(t + k\tau), \quad k \in \mathcal{Z}_+, \quad t \in [0, \tau].$$

The periodic Poisson process  $\{X(t), t \in [0, T], T = n\tau\}$  with the known period  $\tau$  can be transformed to a Poisson process on the interval  $[0, \tau]$ , observed independently  $n$  times, using the following notations

$$\begin{aligned} X_j(t) &= X((j-1)\tau + t) - X((j-1)\tau), \quad t \in [0, \tau], \\ \mathbf{X}_j &= \{X_j(t), t \in [0, \tau]\}, \quad j = 1, \dots, n. \end{aligned}$$

Hence, we observe i.i.d. observations of an inhomogeneous Poisson process on a segment of the real line  $\mathbf{X}^n = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$  where  $\mathbf{X}_j = \{X_j(t), t \in [0, \tau]\}$  is a Poisson process on the interval  $[0, \tau]$ . The estimation problem of the intensity function is well developed in various works by several authors. We mention here only two of them [21], [37] and an upcoming book [20]. In this chapter we consider the problem of estimation of the mean function  $\Lambda(t)$ .

For a Poisson process  $X(t)$  with the mean function  $\Lambda(t)$  denoting by  $\pi(t) = X(t) - \Lambda(t)$ ,

$$X(t) = \Lambda(t) + \pi(t),$$

therefore, the estimation problem of the mean function can be seen as an information transmission problem, where the noise  $\pi(t)$ , unlike the classical Gaussian white noise model, depends on the unknown function  $\Lambda(t)$ . Indeed,  $\mathbf{E}\pi(t) = 0$ ,  $\text{Var } \pi(t) = \Lambda(t)$ .

Suppose  $\mathbf{X} = \{X(t), t \in [0, \tau]\}$  is a Poisson process which induces a measure on the Skorokhod space  $(\mathbb{D}, \mathcal{B}(\mathbb{D}), \{\mathbf{P}_\theta, \theta \in \Theta = [\alpha, \beta]\})$  from the given parametric family. Denote by  $\Lambda(\theta, t)$  the mean functions of the corresponding measures. If  $\sup_{\theta \in \Theta} \Lambda(\theta, \tau) < +\infty$  and  $\lambda(\theta, \cdot) > 0$ , then the measures  $\{\mathbf{P}_\theta, \theta \in \Theta = [\alpha, \beta]\}$  are equivalent and the *Radon-Nykodim derivative* (the likelihood function) for all  $\theta, \theta_1 \in \Theta$  equals

$$\ln L(\theta, \theta_1, \mathbf{X}) = \int_0^\tau \ln \frac{\lambda(\theta, t)}{\lambda(\theta_1, t)} dX(t) - \int_0^\tau [\lambda(\theta, t) - \lambda(\theta_1, t)] dt.$$

Take  $\lambda(\theta_1, t) \equiv \lambda$ ,

$$\ln L(\theta, \mathbf{X}) = \int_0^\tau [\ln \lambda(\theta, t) - \ln \lambda] dX(t) - \int_0^\tau [\lambda(\theta, t) - \lambda] dt,$$

then, the derivative of the logarithm of the likelihood ratio equals  $(\pi(\theta, t) = X(t) - \Lambda(\theta, t))$

$$\frac{\partial \ln L(\theta, \mathbf{X})}{\partial \theta} = \int_0^\tau \frac{\frac{\partial \lambda(\theta, t)}{\partial \theta}}{\lambda(\theta, t)} dX(t) - \int_0^\tau \frac{\partial \lambda(\theta, t)}{\partial \theta} dt = \int_0^\tau \frac{\frac{\partial \lambda(\theta, t)}{\partial \theta}}{\lambda(\theta, t)} d\pi(\theta, t).$$

Finally, the Fisher information of a Poisson process equals

$$I(\theta) = \mathbf{E}_\theta \left( \frac{\partial \ln L(\theta, \mathbf{X})}{\partial \theta} \right)^2 = \int_0^\tau \left[ \frac{\frac{\partial \lambda(\theta, t)}{\partial \theta}}{\lambda(\theta, t)} \right]^2 \lambda(\theta, t) dt. \quad (1.1)$$

Each function  $\bar{\Lambda}_n(t) = \bar{\Lambda}_n(t, \mathbf{X}^n)$  which is measurable with respect to (w.r.t.) the observations is called an *estimator*. This general definition of an estimator does not depend on the estimating object, but obviously, to have “good” estimators we need to compare it with the unknown object, subject to the estimation. For this reason we need a loss function, which, in our case, is the expectation of the  $L_2$ -norm, the mean integrated squared error (MISE) denoted by  $\mathbf{E}_\Lambda \|\bar{\Lambda} - \Lambda\|^2$ , here  $\|\Lambda\|^2 = \int_0^\tau \Lambda^2(t) dt$ ,  $\Lambda \in L_2[0, \tau]$ , (for the estimators not square-integrable the MISE is supposed  $+\infty$ ) and  $\mathbf{E}_\Lambda$  is the mathematical expectation w.r.t. the measure induced by the Poisson process under the assumption that its true mean function is the given function  $\Lambda(\cdot)$ . A very simple, on the other hand, an asymptotically the best estimator (see below for definitions) is *the empirical mean function* (the EMF)

$$\hat{\Lambda}_n(t) = \frac{1}{n} \sum_{j=1}^n X_j(t).$$

For our subsequent results an important role will play the equality

$$\mathbf{E}_\Lambda \|\sqrt{n}(\hat{\Lambda}_n - \Lambda)\|^2 = \int_0^\tau \Lambda(t) dt. \quad (1.2)$$

This means that the EMF tends to the unknown mean function in expected square norm, that the rate of convergence is the classical  $\sqrt{n}$  rate, as in parametric problems, and the asymptotic value of the MISE (which is, in fact, non-asymptotic) is the right-hand side of the equation. The first question that arises is the possibility of construction of estimators with the rate of convergence higher than  $\sqrt{n}$  and for the highest (*optimal*) rate the minimal (attainable) asymptotic MISE. The answer of the first part of this question is negative and the minimal (attainable) asymptotic MISE is that of the EMF. The following theorem of Kutoyants ([20]) states these facts.

Denote by  $\Sigma(R)$  the set of absolutely continuous mean functions centered around given (arbitrary) absolutely continuous mean function  $\Lambda_*$ , with intensity functions satisfying

$$\Sigma(R) = \left\{ \Lambda, \int_0^\tau (\lambda(t) - \lambda_*(t))^2 dt \leq R, \Lambda_*(0) = \Lambda(0) = 0, \Lambda_*(\tau) = \Lambda(\tau) \right\}.$$

**Theorem 1.1.** *In the  $L_2$  estimation problem of the mean function the optimal rate of convergence is  $\sqrt{n}$  and the minimal asymptotic MISE is that of the empirical mean function, uniformly over the set  $\Sigma(R)$*

$$\lim_{n \rightarrow +\infty} \inf_{\tilde{\Lambda}_n} \sup_{\Lambda \in \Sigma(R)} \left( \mathbf{E}_\Lambda \|\sqrt{n}(\tilde{\Lambda}_n - \Lambda)\|^2 - \int_0^\tau \Lambda(t) dt \right) = 0, \quad (1.3)$$

where  $\inf_{\tilde{\Lambda}_n}$  is taken over all possible estimators.

The theorem in [20] is stated slightly differently (for shrinking vicinities). The proof is given in the Appendix.

**Remark 1.1.** *This result is true for sets other than  $\Sigma(R)$  too (see in the Appendix). For example, we can take an  $L_2$  ball centered around  $\Lambda^*$  or take higher order derivatives in the definition of  $\Sigma(R)$*

$$\int_0^\tau (\lambda^{(p)}(t) - \lambda_*^{(p)}(t))^2 dt \leq R, \quad p \in \mathcal{N}.$$

*In fact, if we take as a set only the center of the above set  $\{\Lambda_*\}$ , the result is not true for arbitrary  $\Lambda_*$  (there is a counterexample of an estimator with smaller asymptotic MISE than that of the EMF, it is called the Hodges estimator ([20]).) The set has to be sufficiently “rich” ([10]), that is, to contain certain (exponential) functions. The result is true for all sets containing these functions. We will characterize these functions during the proof.  $\diamond$*

### 1.3 First Order Efficiency

In this section we are going to construct a class of estimators that also satisfy the equation (1.3), thus are also asymptotically efficient (as the EMF). Consider the *kernel-type estimators*, defined as the convolution of the EMF

$$\tilde{\Lambda}_n(t) = \int_0^\tau K_n(s-t) \hat{\Lambda}_n(s) ds,$$



where the kernels  $K_n(\cdot)$  are non-negative, integrable, normalized functions on  $[-\frac{\tau}{2}, \frac{\tau}{2}]$ ,

$$K_n(u) \geq 0, \quad \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} K_n(u) du = 1, \quad n \in \mathcal{N},$$

that are symmetric and we continue them  $\tau$  periodically on the whole real line  $\mathbf{R}$

$$K_n(u) = K_n(-u), \quad K_n(u) = K_n(u + k\tau), \quad u \in \left[-\frac{\tau}{2}, \frac{\tau}{2}\right], \quad k \in \mathcal{Z}. \quad (1.4)$$

Consider the trigonometric basis, which is a complete, orthonormal basis in  $L_2[0, \tau]$

$$\phi_1(t) = \sqrt{\frac{1}{\tau}}, \quad \phi_{2l}(t) = \sqrt{\frac{2}{\tau}} \cos \frac{2\pi l}{\tau} t, \quad \phi_{2l+1}(t) = \sqrt{\frac{2}{\tau}} \sin \frac{2\pi l}{\tau} t, \quad l \in \mathcal{N}.$$

A simple calculation of the Fourier coefficients of the kernel-type estimator (1.5) w.r.t. this basis gives us

$$\tilde{\Lambda}_{1,n} = \hat{\Lambda}_{1,n}, \quad \tilde{\Lambda}_{2l,n} = \sqrt{\frac{\tau}{2}} K_{2l,n} \hat{\Lambda}_{2l,n}, \quad \tilde{\Lambda}_{2l+1,n} = \sqrt{\frac{\tau}{2}} K_{2l,n} \hat{\Lambda}_{2l+1,n}, \quad l \in \mathcal{N},$$

where

$$\tilde{\Lambda}_{l,n} = \int_0^\tau \tilde{\Lambda}_n(t) \phi_l(t) dt, \quad \hat{\Lambda}_{l,n} = \int_0^\tau \hat{\Lambda}_n(t) \phi_l(t) dt, \quad K_{l,n} = \int_0^\tau K_n(t) \phi_l(t) dt.$$

Indeed,

$$\begin{aligned} \tilde{\Lambda}_{l,n} &= \int_0^\tau \tilde{\Lambda}_n(t) \phi_l(t) dt = \frac{1}{n} \sum_{j=1}^n \int_0^\tau X_j(s) \left( \int_0^\tau K_n(s-t) \phi_l(t) dt \right) ds \\ &= \frac{1}{n} \sum_{j=1}^n \int_0^\tau X_j(s) \left( \int_{s-\tau}^s K_n(u) \phi_l(s-u) du \right) ds. \end{aligned}$$

We calculate separately the even and odd Fourier coefficients

$$\begin{aligned} \tilde{\Lambda}_{2l+1,n} &= \sqrt{\frac{2}{\tau}} \frac{1}{n} \sum_{j=1}^n \int_0^\tau X_j(s) \left( \int_{s-\tau}^s K_n(u) \sin \frac{2\pi l}{\tau} (s-u) du \right) ds \\ &= \sqrt{\frac{2}{\tau}} \frac{1}{n} \sum_{j=1}^n \int_0^\tau X_j(s) \sin \frac{2\pi l}{\tau} s \left( \int_{s-\tau}^s K_n(u) \cos \frac{2\pi l}{\tau} u du \right) ds \\ &\quad - \sqrt{\frac{2}{\tau}} \frac{1}{n} \sum_{j=1}^n \int_0^\tau X_j(s) \cos \frac{2\pi l}{\tau} s \left( \int_{s-\tau}^s K_n(u) \sin \frac{2\pi l}{\tau} u du \right) ds. \end{aligned}$$

Since

$$\begin{aligned}\int_{s-\tau}^s K_n(u) \sin \frac{2\pi l}{\tau} u du &= \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} K_n(u) \sin \frac{2\pi l}{\tau} u du = 0, \\ \int_{s-\tau}^s K_n(u) \cos \frac{2\pi l}{\tau} u du &= \int_0^\tau K_n(u) \cos \frac{2\pi l}{\tau} u du = \sqrt{\frac{\tau}{2}} K_{2l,n},\end{aligned}$$

then (the second one can be proved in the same way)

$$\tilde{\Lambda}_{2l+1,n} = \sqrt{\frac{\tau}{2}} K_{2l,n} \cdot \hat{\Lambda}_{2l+1,n}, \quad \tilde{\Lambda}_{2l,n} = \sqrt{\frac{\tau}{2}} K_{2l,n} \cdot \hat{\Lambda}_{2l,n}.$$

The kernel-type estimators are linear estimators and we are going to prove asymptotic efficiency of these estimators for appropriately chosen kernels. To anticipate the theory that we will develop in the next section, we modify these estimators to fit into the set  $\Sigma(R)$ . This set is centered around the (known) mean function  $\Lambda_*$ , hence we will consider the following kernel-type estimators (we freely use  $\Lambda_*$  in the definition of the estimators since it is known)

$$\tilde{\Lambda}_n(t) = \int_0^\tau K_n(s-t)(\hat{\Lambda}_n(s) - \Lambda_*(s)) ds + \Lambda_*(t). \quad (1.5)$$

Then, their Fourier coefficients will be

$$\begin{aligned}\tilde{\Lambda}_{1,n} &= \hat{\Lambda}_{1,n}, \quad \tilde{\Lambda}_{2l,n} = \sqrt{\frac{\tau}{2}} K_{2l,n}(\hat{\Lambda}_{2l,n} - \Lambda_{2l}^*) + \Lambda_{2l}^*, \\ \tilde{\Lambda}_{2l+1,n} &= \sqrt{\frac{\tau}{2}} K_{2l,n}(\hat{\Lambda}_{2l+1,n} - \Lambda_{2l+1}^*) + \Lambda_{2l+1}^*, \quad l \in \mathcal{N},\end{aligned} \quad (1.6)$$

with

$$\Lambda_l = \int_0^\tau \Lambda(t) \phi_l(t) dt, \quad \Lambda_l^* = \int_0^\tau \Lambda_*(t) \phi_l(t) dt.$$

We are proceeding with the calculation of SMISE for the introduced estimators. Using the Parseval equality with the (1.6), and denoting  $\sigma_{l,n}^2 = \mathbf{E}|\hat{\Lambda}_{l,n} - \Lambda_l|^2$ , we get

$$\begin{aligned}\mathbf{E}_\Lambda \|\sqrt{n}(\tilde{\Lambda}_n - \Lambda)\|^2 - \mathbf{E}_\Lambda \|\sqrt{n}(\hat{\Lambda}_n - \Lambda)\|^2 &= \sum_{l=1}^{+\infty} \left( \frac{\tau}{2} K_{2l,n}^2 - 1 \right) n(\sigma_{2l,n}^2 + \sigma_{2l,n}^2) \\ &\quad + \sum_{l=1}^{+\infty} n \left| \sqrt{\frac{\tau}{2}} K_{2l,n} - 1 \right|^2 [(\Lambda_{2l} - \Lambda_{2l}^*)^2 + (\Lambda_{2l+1} - \Lambda_{2l+1}^*)^2].\end{aligned} \quad (1.7)$$

Using the fact that (see the equation (1.2))

$$\mathbf{E}_\Lambda \|\hat{\Lambda}_n - \Lambda\|^2 = \sum_{l=1} \sigma_{l,n}^2 = \frac{1}{n} \int_0^\tau \Lambda(t) dt \leq \frac{\tau}{n} \Lambda^*(\tau),$$

we obtain from (1.7) the following upper bound

$$\begin{aligned} \mathbf{E}_\Lambda \|\sqrt{n}(\tilde{\Lambda}_n - \Lambda)\|^2 - \int_0^\tau \Lambda(t) dt &\leq \sup_{l \geq 1} \left( \frac{\tau}{2} K_{2l,n}^2 - 1 \right) \tau \Lambda^*(\tau) + \\ &+ n \sup_{l \geq 1} \left| \sqrt{\frac{\tau}{2}} K_{2l,n} - 1 \right|^2 \sum_{l=1}^{+\infty} [(\Lambda_{2l} - \Lambda_{2l}^*)^2 + (\Lambda_{2l+1} - \Lambda_{2l+1}^*)^2]. \end{aligned}$$

Since  $\|\Lambda\| \leq \|\Lambda - \Lambda_*\| + \|\Lambda_*\| \leq \sqrt{R} + \|\Lambda_*\|$  and using the fact  $\sqrt{\frac{\tau}{2}} K_{2l,n} \leq 1$ , we get

$$1 - \frac{\tau}{2} K_{2l,n}^2 \leq 2 \left( 1 - \sqrt{\frac{\tau}{2}} K_{2l,n} \right),$$

and also because of the inequality

$$\sum_{l=1}^{+\infty} [(\Lambda_{2l} - \Lambda_{2l}^*)^2 + (\Lambda_{2l+1} - \Lambda_{2l+1}^*)^2] \leq R,$$

we state the first result on the construction of asymptotically efficient estimators (see the Remark 1.1, since we maximize here over the ball  $\mathcal{B}(R)$ , not the set  $\Sigma(R)$  as it is stated in the Theorem 1.1).

**Proposition 1.1.** *For the kernel-type estimator (1.5) with a kernel satisfying (1.4) and the condition*

$$n \sup_{l \geq 1} \left| \sqrt{\frac{\tau}{2}} K_{2l,n} - 1 \right|^2 \longrightarrow 0,$$

as  $n \rightarrow +\infty$ , over a  $\Lambda_*$  centered  $\mathcal{B}(R) = \{\Lambda : \int_0^\tau (\Lambda(t) - \Lambda_*(t))^2 dt \leq R\}$  ball of mean functions in  $L_2$ , the following equality holds

$$\lim_{n \rightarrow +\infty} \sup_{\Lambda \in \mathcal{B}(R)} \left( \mathbf{E}_\Lambda \|\sqrt{n}(\tilde{\Lambda}_n - \Lambda)\|^2 - \int_0^\tau \Lambda(t) dt \right) = 0.$$

In fact, it is a variation of the EMF and, in some sense, is artificially created. We would like to enlarge the class of asymptotically efficient estimators, in order to have much more practical estimators. From the second term in the right-hand side of (1.7), we see that imposing additional regularity conditions on the unknown mean function (we replace the ball  $\mathcal{B}(R)$  in  $L_2$  by the compact set  $\Sigma(R)$  which is called a *Sobolev ellipsoid*), we can relax the condition of uniform convergence of

Fourier coefficients of the kernel to 1. And one of the reasons of doing this is that we would like also to describe a class of asymptotically efficient estimators not in terms of the Fourier coefficients of the kernel, but to give the kernel function explicitly (not by its Fourier series). The second result on construction of asymptotically efficient estimators is given by the following proposition.

**Proposition 1.2.** *For the kernel-type estimator (1.5) with a kernel satisfying (1.4) and the condition*

$$n \sup_{l \geq 1} \left| \frac{\sqrt{\frac{\tau}{2}} K_{2l,n} - 1}{\frac{2\pi l}{\tau}} \right|^2 \rightarrow 0,$$

as  $n \rightarrow +\infty$ , the following equality holds

$$\lim_{n \rightarrow +\infty} \sup_{\Lambda \in \Sigma(R)} \left( \mathbf{E}_\Lambda \|\sqrt{n}(\tilde{\Lambda}_n - \Lambda)\|^2 - \int_0^\tau \Lambda(t) dt \right) = 0,$$

where the set  $\Sigma(R)$  is defined in the Theorem 1.1.

*Proof.* Using the equality (1.7) and denoting  $\tilde{\sigma}_{l,n}^2 = n(\sigma_{2l,n}^2 + \sigma_{2l+1,n}^2) \left(\frac{2\pi l}{\tau}\right)^2$ , we can write

$$\begin{aligned} \mathbf{E}_\Lambda \|\sqrt{n}(\tilde{\Lambda}_n - \Lambda)\|^2 - \mathbf{E}_\Lambda \|\sqrt{n}(\hat{\Lambda}_n - \Lambda)\|^2 &= \sum_{l=1}^{+\infty} \frac{\left(\frac{\tau}{2} K_{2l,n}^2 - 1\right)}{\left(\frac{2\pi l}{\tau}\right)^2} \tilde{\sigma}_{l,n}^2 \\ &+ n \sup_{l \geq 1} \frac{\left(\frac{\tau}{2} K_{2l,n} - 1\right)^2}{\left(\frac{2\pi l}{\tau}\right)^2} \sum_{l=1}^{+\infty} \left[ \left(\frac{2\pi l}{\tau}\right)^2 (\Lambda_{2l} - \Lambda_{2l}^*)^2 + (\Lambda_{2l+1} - \Lambda_{2l+1}^*)^2 \right]. \end{aligned} \quad (1.8)$$

Simple calculations give us ([6])  $\sigma_{2l,n}^2 + \sigma_{2l+1,n}^2 = \frac{2}{n} \left(\frac{\tau}{2\pi l}\right)^2 \left(\frac{2}{\tau} \Lambda(\tau) - \sqrt{\frac{2}{\tau}} \lambda_{2l}\right)$ , where  $\lambda_l = \int_0^\tau \lambda(t) \phi_l(t) dt$ . Whence  $\tilde{\sigma}_{l,n}^2 = \frac{4}{\tau} \Lambda(\tau) - 2\sqrt{\frac{2}{\tau}} \lambda_{2l}$ .

The mean function  $\Lambda$  belongs to the set  $\Sigma(R)$  if and only if its Fourier coefficients satisfy the inequality (see the proof of the Proposition 1.4)

$$\sum_{l=1}^{+\infty} \left[ \left(\frac{2\pi l}{\tau}\right)^2 (\Lambda_{2l} - \Lambda_{2l}^*)^2 + (\Lambda_{2l+1} - \Lambda_{2l+1}^*)^2 \right] \leq R,$$

therefore, we can write

$$\begin{aligned} |\mathbf{E}_\Lambda \|\sqrt{n}(\tilde{\Lambda}_n - \Lambda)\|^2 - \mathbf{E}_\Lambda \|\sqrt{n}(\hat{\Lambda}_n - \Lambda)\|^2| &\leq \\ \frac{4}{\tau} \Lambda(\tau) \sum_{l=1}^{+\infty} \frac{1 - \frac{\tau}{2} K_{2l,n}^2}{\left(\frac{2\pi l}{\tau}\right)^2} - 2\sqrt{\frac{2}{\tau}} \sum_{l=1}^{+\infty} \frac{1 - \frac{\tau}{2} K_{2l,n}^2}{\left(\frac{2\pi l}{\tau}\right)^2} \lambda_{2l} &+ n \sup_{l \geq 1} \frac{\left(\sqrt{\frac{\tau}{2}} K_{2l,n} - 1\right)^2}{\left(\frac{2\pi l}{\tau}\right)^2} R. \end{aligned}$$

Denote  $h_n = \sup_{l \geq 1} \frac{1 - \sqrt{\frac{\tau}{2}} K_{2l,n}}{\frac{2\pi l}{\tau}}$ . Evidently,  $\frac{1 - \sqrt{\frac{\tau}{2}} K_{2l,n}}{\frac{2\pi l}{\tau}} \leq h_n$ ,  $l \geq 1$ . Consider the set of all sequences that satisfy this inequality

$$\mathcal{K} = \left\{ \bar{K}_{2l,n}, \left| \sqrt{\frac{\tau}{2}} \bar{K}_{2l,n} \right| \leq 1, \frac{1 - \sqrt{\frac{\tau}{2}} \bar{K}_{2l,n}}{\frac{2\pi l}{\tau}} \leq h_n \right\}.$$

Now, find the minimal sequence on the set  $\mathcal{K}$

$$K_{2l,n}^0 = \arg \min_{\mathcal{K}} |\bar{K}_{2l,n}| = \sqrt{\frac{2}{\tau}} \left( 1 - h_n \frac{2\pi l}{\tau} \right)_+.$$

Obviously, passing from  $K_{2l,n}$  to  $K_{2l,n}^0$  will increase the right-hand side of (1.8). Hence

$$\begin{aligned} & |\mathbf{E}_\Lambda \|\sqrt{n}(\tilde{\Lambda}_n - \Lambda)\|^2 - \mathbf{E}_\Lambda \|\sqrt{n}(\hat{\Lambda}_n - \Lambda)\|^2| \leq \\ & \frac{4}{\tau} \Lambda(\tau) \sum_{l=1}^{+\infty} \frac{1 - \frac{\tau}{2} (K_{2l,n}^0)^2}{\left(\frac{2\pi l}{\tau}\right)^2} - 2\sqrt{\frac{2}{\tau}} \sum_{l=1}^{+\infty} \frac{1 - \frac{\tau}{2} (K_{2l,n}^0)^2}{\left(\frac{2\pi l}{\tau}\right)^2} \lambda_{2l} + nh_n R. \end{aligned} \quad (1.9)$$

The third term of the right-hand side tends to zero, as  $n \rightarrow +\infty$ , because of the condition of the theorem. For the second term, use the Cauchy-Schwarz inequality

$$\begin{aligned} & \sum_{l=1}^{+\infty} \frac{1 - \frac{\tau}{2} (K_{2l,n}^0)^2}{\left(\frac{2\pi l}{\tau}\right)^2} \lambda_{2l} \leq 2 \sum_{l=1}^{+\infty} \frac{1 - \frac{\tau}{2} K_{2l,n}^0}{\left(\frac{2\pi l}{\tau}\right)^2} \lambda_{2l} \leq 2h_n \sum_{l=1}^{+\infty} \frac{\tau}{2\pi l} \lambda_{2l} \\ & \leq 2h_n \left( \sum_{l=1}^{+\infty} \left(\frac{\tau}{2\pi l}\right)^2 \right)^{\frac{1}{2}} \left( \sum_{l=1}^{+\infty} \lambda_{2l}^2 \right)^{\frac{1}{2}} \leq 2h_n \left( \sum_{l=1}^{+\infty} \left(\frac{\tau}{2\pi l}\right)^2 \right)^{\frac{1}{2}} \|\lambda\| \\ & \leq 2h_n \left( \sum_{l=1}^{+\infty} \left(\frac{\tau}{2\pi l}\right)^2 \right)^{\frac{1}{2}} (\sqrt{R} + \|\lambda_*\|). \end{aligned}$$

Here we have used the fact that  $\sum_{l=1}^{+\infty} \lambda_l^2 = \int_0^\tau \lambda^2(t) dt$  and that  $\Lambda \in \Sigma(R)$ . Therefore, the second term in (1.9) also has a numerical sequence as its upper bound which tends to zero.

Consider the first term of (1.9). First,  $\Lambda(\tau) = \Lambda_*(\tau)$ . Then, denoting the entire

part of following sequence by  $N_n = \left\lceil \frac{\tau}{2\pi} \frac{1}{h_n} \right\rceil$ , we find

$$\begin{aligned} \sum_{l=1}^{+\infty} \frac{1 - \frac{\tau}{2}(K_{2l,n}^0)^2}{\left(\frac{2\pi l}{\tau}\right)^2} &= \sum_{l=1}^{N_n} \frac{1 - \frac{\tau}{2}(K_{2l,n}^0)^2}{\left(\frac{2\pi l}{\tau}\right)^2} + \sum_{l=N_n+1}^{+\infty} \left(\frac{\tau}{2\pi l}\right)^2 = \\ \sum_{l=1}^{N_n} \frac{2h_n \frac{2\pi l}{\tau} - h_n^2 \left(\frac{2\pi l}{\tau}\right)^2}{\left(\frac{2\pi l}{\tau}\right)^2} + \sum_{l=N_n+1}^{+\infty} \left(\frac{\tau}{2\pi l}\right)^2 &= \\ \frac{h_n \tau}{\pi} \sum_{l=1}^{N_n} \frac{1}{l} - h_n^2 N_n + \sum_{l=N_n+1}^{+\infty} \left(\frac{\tau}{2\pi l}\right)^2. \end{aligned}$$

Using the fact that the following limit exists (it is a positive number smaller than 1 and is called the Euler-Mascheroni constant)

$$\gamma = \lim_{n \rightarrow +\infty} \gamma_n, \quad \gamma_n = \sum_{l=1}^n \frac{1}{l} - \ln n,$$

and the fact that  $\sum_{l=N_n+1}^{+\infty} \frac{1}{l^2} \rightarrow 0$ , as  $n \rightarrow +\infty$ , we get

$$\sum_{l=1}^{+\infty} \frac{1 - \frac{\tau}{2}(K_{2l,n}^0)^2}{\left(\frac{2\pi l}{\tau}\right)^2} = \frac{h_n \tau}{\pi} (\gamma_{N_n} + \ln N_n) - h_n^2 N_n + \sum_{l=N_n+1}^{+\infty} \left(\frac{\tau}{2\pi l}\right)^2,$$

which tends to zero, since  $h_n^2 n \rightarrow 0$ , as  $n \rightarrow +\infty$ . Therefore, the first term of (1.9) also tends to zero.  $\blacksquare$

**Remark 1.2.** Suppose that  $K(u)$ ,  $u \in \left[-\frac{\tau}{2}, \frac{\tau}{2}\right]$  is a non-negative, integrable, normalized function

$$K(u) \geq 0, \quad u \in \left[-\frac{\tau}{2}, \frac{\tau}{2}\right], \quad \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} K(u) du = 1,$$

which is symmetric and we continue it  $\tau$  periodically on the whole real line

$$K(u) = K(-u), \quad K(u) = K(u + k\tau), \quad u \in \left[-\frac{\tau}{2}, \frac{\tau}{2}\right], \quad k \in \mathcal{Z}.$$

Let the positive sequence  $h_n \leq 1$  be such that  $h_n^2 n \rightarrow 0$ ,  $n \rightarrow +\infty$ . Then, the kernels

$$K_n(u) = \frac{1}{h_n} K\left(\frac{u}{h_n}\right) \mathbb{I}\left\{|u| \leq \frac{\tau}{2} h_n\right\} \quad (1.10)$$

satisfy (1.4) and the condition of the proposition 1.2.

Indeed, (1.4) is obvious and

$$\begin{aligned}\sqrt{\frac{\tau}{2}}K_{2l,n} - 1 &= \frac{1}{h_n} \int_0^\tau K\left(\frac{u}{h_n}\right) \mathbb{I}\left\{|u| \leq \frac{\tau}{2}h_n\right\} \cos\left(\frac{2\pi l}{\tau}u\right) du - 1 = \\ &= \int_{-\frac{\tau}{2h_n}}^{\frac{\tau}{2h_n}} K(t) \cos\left(\frac{2\pi l}{\tau}h_n t\right) \mathbb{I}\left\{|t| \leq \frac{\tau}{2}\right\} dt - 1 = \\ &= \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} K(t) \left[\cos\left(\frac{2\pi l}{\tau}h_n t\right) - \cos(0)\right] dt.\end{aligned}$$

Since the trigonometric functions satisfy the Lipschitz condition with the constant 1, then

$$\left|\sqrt{\frac{\tau}{2}}K_{2l,n} - 1\right| \leq \frac{2\pi l}{\tau}h_n\tau \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} K(t)dt.$$

Hence, as  $n \rightarrow +\infty$ ,

$$n \sup_{l \geq 1} \left| \frac{\sqrt{\frac{\tau}{2}}K_{2l,n} - 1}{\frac{2\pi l}{\tau}} \right| \leq nh_n\tau \rightarrow 0,$$

which ensures that the kernel type estimator (1.5) with the kernel (1.10) is asymptotically efficient over the set  $\Sigma(R)$ .  $\diamond$

## 1.4 Second Order Efficiency

As it was mentioned in the previous section to establish first order efficiency we do not need any regularity conditions. Now, imposing regularity conditions on the unknown mean function we compare second term of the expansion of the mean integrated square error, which allows us to compare first order asymptotically efficient estimators. An estimator, for which the asymptotic error of the second term is minimal, is called second order asymptotically efficient estimator (see Proposition 1.4). Below, we establish the lower bound for the second term of the expansion of the mean integrated square error, find asymptotic error explicitly and construct a second order asymptotically efficient estimator.

### 1.4.1 Presentation of results

We consider here an ellipsoid centered not around a given mean function, but centered around the null function. For a given integer  $m > 1$  consider the following

set of non-decreasing, positive functions on  $[0, \tau]$  such that their  $(m-1)$ th derivative is absolutely continuous and

$$\mathcal{F}_m^{per}(R, S) = \left\{ \Lambda(\cdot) : \int_0^\tau [\Lambda^{(m)}(t)]^2 dt \leq R, \Lambda(0) = 0, \Lambda(\tau) = S \right\}, \quad m > 1, \quad (1.11)$$

where  $R > 0, S > 0$  are given constants. Periodicity means that the intensity function  $\lambda(\cdot)$  is periodic (hence the equality of its values and the values of its derivatives on the endpoints of the interval  $[0, \tau]$ ). Introduce as well

$$\Pi = \Pi_m(R, S) = (2m - 1)R \left( \frac{S}{\pi R} \frac{m}{(2m - 1)(m - 1)} \right)^{\frac{2m}{2m-1}}. \quad (1.12)$$

**Proposition 1.3.** *Consider Poisson observations  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ . Then, for all estimators  $\bar{\Lambda}_n(t)$  of the mean function  $\Lambda(t)$ , following lower bound holds*

$$\liminf_{n \rightarrow +\infty} \sup_{\Lambda \in \mathcal{F}_m(R, S)} n^{\frac{2m}{2m-1}} \left( \int_0^\tau \mathbf{E}_\Lambda(\bar{\Lambda}_n(t) - \Lambda(t))^2 dt - \frac{1}{n} \int_0^\tau \Lambda(t) dt \right) \geq -\Pi.$$

In the next proposition we propose an estimator which attains this lower bound, thus we will prove that this lower bound is sharp. Introduce

$$\Lambda_n^*(t) = \hat{\Lambda}_{0,n} \phi_0(t) + \sum_{l=1}^{N_n} \tilde{K}_{l,n} \hat{\Lambda}_{l,n} \phi_l(t),$$

where  $\{\phi_l\}_{l=0}^{+\infty}$  is the trigonometric cosine basis in  $L_2[0, \tau]$ ,  $\hat{\Lambda}_{l,n}$  are the Fourier coefficients of the empirical mean function with respect to this basis and

$$\tilde{K}_{l,n} = \left( 1 - \left| \frac{\pi l}{\tau} \right|^m \alpha_n^* \right)_+, \quad \alpha_n^* = \left[ \frac{S}{nR} \frac{\tau}{\pi} \frac{m}{(2m-1)(m-1)} \right]^{\frac{m}{2m-1}},$$

$$N_n = \frac{\tau}{\pi} (\alpha_n^*)^{-\frac{1}{m}} \approx C n^{\frac{1}{2m-1}}, \quad x_+ = \max(x, 0), \quad x \in \mathbf{R}.$$

The next proposition states

**Proposition 1.4.** *The estimator  $\Lambda_n^*(t)$  attains the lower bound described above, that is,*

$$\lim_{n \rightarrow +\infty} \sup_{\Lambda \in \mathcal{F}_m(R, S)} n^{\frac{2m}{2m-1}} \left( \int_0^\tau \mathbf{E}_\Lambda(\Lambda_n^*(t) - \Lambda(t))^2 dt - \frac{1}{n} \int_0^\tau \Lambda(t) dt \right) = -\Pi.$$



### 1.4.2 Proof of Proposition 1.4

Consider the  $L_2[0, \tau]$  Hilbert space. Evidently,  $\mathcal{F}_m^{per}(R, S) \subset L_2[0, \tau]$ . The main idea of the proof is to replace the estimation problem of the infinite-dimensional (continuum) mean function by the estimation problem of infinite-dimensional but countable vector of its Fourier coefficients. Recall that the space  $L_2[0, \tau]$  is isomorphic to the space

$$\ell_2 = \left\{ \theta = (\theta_l)_{l=0}^{+\infty} : \sum_{l=0}^{+\infty} \theta_l^2 < +\infty \right\}, \quad \|\theta\| = \left( \sum_{l=0}^{+\infty} \theta_l^2 \right)^{\frac{1}{2}}.$$

Our first goal is to describe the set  $\Theta \subset \ell_2$  of Fourier coefficients of the functions from the set  $\mathcal{F}_m^{per}(R, S)$ .

Consider a complete, orthonormal system in the space  $L_2[0, \tau]$ ,

$$\phi_0(t) = \sqrt{\frac{1}{\tau}}, \quad \phi_l(t) = \sqrt{\frac{2}{\tau}} \cos \frac{\pi l}{\tau} t, \quad l \in \mathcal{N}. \quad (1.13)$$

For the first order efficiency we used the trigonometric basis with both sine and cosine functions present, here we took only the cosine basis to avoid technical difficulties and, as a result, the proposed estimator does not depend on the center of the ellipsoid (here it is present as the quantity  $S$ ). Each function  $f \in L_2[0, \tau]$  is a  $L_2$ -limit of its Fourier series

$$f(t) = \sum_{l=0}^{+\infty} \theta_l \phi_l(t), \quad \theta_l = \int_0^\tau f(t) \phi_l(t) dt.$$

Suppose that

$$\Lambda_l = \int_0^\tau \Lambda(t) \phi_l(t) dt, \quad \lambda_l = \int_0^\tau \lambda(t) \phi_l(t) dt.$$

Then

**Lemma 1.1.** *The mean function  $\Lambda$  belongs to the set  $\mathcal{F}_m^{per}(R, S)$  if and only if its Fourier coefficients w.r.t. the cosine trigonometric basis satisfy*

$$\sum_{l=1}^{+\infty} \left( \frac{\pi l}{\tau} \right)^{2m} \Lambda_l^2 \leq R, \quad \Lambda(\tau) = S, \quad (1.14)$$

or, the Fourier coefficients of its intensity function satisfy

$$\sum_{l=1}^{+\infty} \left( \frac{\pi l}{\tau} \right)^{2(m-1)} \lambda_l^2 \leq R, \quad \Lambda(\tau) = S. \quad (1.15)$$

For the proof see, for example, [40] (lemma A.3). To introduce the estimator denote the Fourier coefficients of the empirical mean function by

$$\hat{\Lambda}_{l,n} = \int_0^\tau \hat{\Lambda}_n(t) \phi_l(t) dt, \quad l \in \mathcal{Z}_+, \quad \hat{\Lambda}_n(t) = \frac{1}{n} \sum_{j=1}^n X_j(t).$$

Consider the estimator

$$\tilde{\Lambda}_n(t) = \sum_{l=0}^{+\infty} \tilde{\Lambda}_{l,n} \phi_l(t), \quad \tilde{\Lambda}_{l,n} = K_{l,n} \hat{\Lambda}_{l,n}.$$

Here  $K_{l,n}$  are some numbers. Without loss of generality we can take  $K_{0,n} = 1$ , that is  $\tilde{\Lambda}_{0,n} = \hat{\Lambda}_{0,n}$ . In this case, using the Parseval's equality, we get

$$\mathbf{E}_\Lambda \|\tilde{\Lambda}_n - \Lambda\|^2 - \mathbf{E}_\Lambda \|\hat{\Lambda}_n - \Lambda\|^2 = \sum_{l=1}^{+\infty} (K_{l,n}^2 - 1) \sigma_{l,n}^2 + \sum_{l=1}^{+\infty} |K_{l,n} - 1|^2 \Lambda_l^2. \quad (1.16)$$

Here  $\sigma_{l,n}^2 = \mathbf{E}_\Lambda (\hat{\Lambda}_{l,n} - \Lambda_l)^2$ . To compute this quantity, introduce the notation

$$\pi_j(t) = X_j(t) - \Lambda(t).$$

In the sequel, we are going to use the following property of stochastic integral (see, for example, [18])

$$\mathbf{E}_\Lambda \left[ \int_0^\tau f(t) d\pi_j(t) \int_0^\tau g(t) d\pi_j(t) \right] = \int_0^\tau f(t) g(t) d\Lambda(t), \quad f, g \in L_2[0, \tau].$$

Further, in view of the integration by parts, we have

$$\hat{\Lambda}_{l,n} - \Lambda_l = \frac{1}{n} \sum_{j=1}^n \int_0^\tau \pi_j(t) \phi_l(t) dt = \frac{1}{n} \sum_{j=1}^n \int_0^\tau \left( \int_t^\tau \phi_l(s) ds \right) d\pi_j(t).$$

which entails that

$$\sigma_{l,n}^2 = \mathbf{E}_\Lambda |\hat{\Lambda}_{l,n} - \Lambda_l|^2 = \frac{1}{n} \int_0^\tau \left( \int_t^\tau \phi_l(s) ds \right)^2 d\Lambda(t).$$

Simple algebra yields

$$\sigma_{l,n}^2 = \frac{1}{n} \left( \frac{\tau}{\pi l} \right)^2 \left[ \Lambda(\tau) - \frac{2}{\tau} \int_0^\tau \cos \left( \frac{2\pi l}{\tau} t \right) \lambda(t) dt \right].$$

Combining with (1.16), this leads to

$$\begin{aligned} \mathbf{E}_\Lambda \|\tilde{\Lambda}_n - \Lambda\|^2 - \mathbf{E}_\Lambda \|\hat{\Lambda}_n - \Lambda\|^2 &= \frac{S}{n} \sum_{l=1}^{+\infty} \left(\frac{\tau}{\pi l}\right)^2 (K_{l,n}^2 - 1) \\ &+ \sum_{l=1}^{+\infty} (K_{l,n} - 1)^2 \Lambda_l^2 + \frac{1}{n} \sqrt{\frac{2}{\tau}} \sum_{l=1}^{+\infty} \left(\frac{\tau}{\pi l}\right)^2 (1 - K_{l,n}^2) \lambda_{2l}. \end{aligned} \quad (1.17)$$

For the third term in the right-hand side we have

$$\begin{aligned} &\left| \frac{1}{n} \sqrt{\frac{2}{\tau}} \sum_{l=1}^{+\infty} \left(\frac{\tau}{\pi l}\right)^2 (1 - K_{l,n}^2) \lambda_{2l} \right| \leq \\ &\leq \frac{1}{n} \sqrt{\frac{2}{\tau}} \max_l \frac{|1 - K_{l,n}^2|}{\left(\frac{\pi l}{\tau}\right)^m} \sum_{l=1}^{+\infty} \left(\frac{\pi l}{\tau}\right)^{m-1} \lambda_{2l} \left(\frac{\pi l}{\tau}\right)^{-1} \\ &\leq \frac{1}{n} \sqrt{\frac{2}{\tau}} \max_l \frac{|1 - K_{l,n}^2|}{\left(\frac{\pi l}{\tau}\right)^m} \left( \sum_{l=1}^{+\infty} \left(\frac{\pi l}{\tau}\right)^{2(m-1)} \lambda_{2l}^2 \right)^{\frac{1}{2}} \left( \sum_{l=1}^{+\infty} \left(\frac{\pi l}{\tau}\right)^{-2} \right)^{\frac{1}{2}}. \end{aligned}$$

Using (1.15) from the Lemma 1.1 we obtain

$$\left( \sum_{l=1}^{+\infty} \left(\frac{\pi l}{\tau}\right)^{2(m-1)} \lambda_{2l}^2 \right)^{\frac{1}{2}} \leq \sqrt{R}.$$

Hence

$$\left| \frac{1}{n} \sqrt{\frac{2}{\tau}} \sum_{l=1}^{+\infty} \left(\frac{\tau}{\pi l}\right)^2 (1 - K_{l,n}^2) \lambda_{2l} \right| \leq \frac{C}{n} \max_l \frac{|1 - |K_{l,n}|^2|}{\left(\frac{\pi l}{\tau}\right)^m}$$

Now, consider the first two terms of the right-hand side of the equation (1.17). Introduce a set of possible kernels (for all  $c_n > 0$ )

$$\mathcal{C}_n = \left\{ K_{l,n} : |K_{l,n} - 1| \leq \left| \frac{\pi l}{\tau} \right|^m c_n \right\}.$$

From (1.14) follows

$$\begin{aligned} \frac{S}{n} \sum_{l=1}^{+\infty} \left(\frac{\tau}{\pi l}\right)^2 (K_{l,n}^2 - 1) + \sum_{l=1}^{+\infty} |K_{l,n} - 1|^2 \Lambda_l^2 &= \frac{S}{n} \sum_{l=1}^{+\infty} \left(\frac{\tau}{\pi l}\right)^2 (K_{l,n}^2 - 1) + \\ &\sum_{l=1}^{+\infty} \frac{|K_{l,n} - 1|^2}{\left(\frac{\pi l}{\tau}\right)^{2m}} \left(\frac{\pi l}{\tau}\right)^{2m} \Lambda_l^2 \leq \frac{S}{n} \sum_{l=1}^{+\infty} \left(\frac{\tau}{\pi l}\right)^2 (K_{l,n}^2 - 1) + c_n^2 R. \end{aligned}$$

Hence, minimizing the later over the set  $\mathcal{C}_n$

$$\tilde{K}_{l,n} = \arg \min_{c_n} |K_{l,n}| = \left(1 - \left|\frac{\pi l}{\tau}\right|^m c_n\right)_+, \quad (1.18)$$

we obtain

$$\begin{aligned} & \sup_{\Lambda \in \mathcal{F}_m^{per}(R,S)} \left( \mathbf{E}_\Lambda \|\tilde{\Lambda}_n - \Lambda\|^2 - \mathbf{E}_\Lambda \|\hat{\Lambda}_n - \Lambda\|^2 \right) \leq \\ & \frac{S}{n} \sum_{l=1}^{+\infty} \left(\frac{\tau}{\pi l}\right)^2 (\tilde{K}_{l,n}^2 - 1) + c_n^2 R + \frac{C}{n} \max_l \frac{|1 - \tilde{K}_{l,n}^2|}{\left(\frac{\pi l}{\tau}\right)^m}. \end{aligned} \quad (1.19)$$

Here  $\tilde{\Lambda}_n(t)$  is the estimator corresponding to the kernel  $\tilde{K}(u)$ . In fact, we have not yet constructed the estimator. We have to specify the sequence of positive numbers  $c_n$  in the definition (1.18). Consider the function

$$H(c_n) = \frac{S}{n} \sum_{l=1}^{+\infty} \left(\frac{\tau}{\pi l}\right)^2 (\tilde{K}_{l,n}^2 - 1) + c_n^2 R$$

and minimize it with respect to the positive sequence  $c_n$ . Introduce as well  $N_n = \frac{\tau}{\pi} c_n^{-\frac{1}{m}}$ . Then

$$H(c_n) = \frac{S}{n} \left[ \sum_{l \leq N_n} \left(\frac{\tau}{\pi l}\right)^2 \left( c_n^2 \left(\frac{\pi l}{\tau}\right)^{2m} - 2c_n \left(\frac{\pi l}{\tau}\right)^m \right) - \sum_{l > N_n} \left(\frac{\tau}{\pi l}\right)^2 \right] + c_n^2 R.$$

To minimize this function consider its derivative

$$H'(c_n) = \frac{S}{n} \sum_{l \leq N_n} \left(\frac{\tau}{\pi l}\right)^2 \left[ 2c_n \left(\frac{\pi l}{\tau}\right)^{2m} - 2 \left(\frac{\pi l}{\tau}\right)^m \right] + 2c_n R = 0. \quad (1.20)$$

Consider such sums ( $\beta \in \mathcal{N}$ )

$$\sum_{l \leq N_n} l^\beta = \sum_{l=1}^{[N_n]} \left(\frac{l}{[N_n]}\right)^\beta [N_n]^\beta = [N_n]^{\beta+1} \sum_{l=1}^{[N_n]} \left(\frac{l}{[N_n]}\right)^\beta \frac{1}{[N_n]},$$

hence, if  $c_n \rightarrow 0$ , as  $n \rightarrow +\infty$ ,

$$\frac{1}{[N_n]^{\beta+1}} \sum_{l \leq N_n} l^\beta \rightarrow \int_0^1 x^\beta dx,$$

that is,

$$\sum_{l \leq N_n} l^\beta = \frac{[N_n]^{\beta+1}}{\beta+1} (1 + o(1)), \quad n \rightarrow +\infty.$$

Using this identity we can transform (1.19) (remembering that  $N_n = \frac{\tau}{\pi} c_n^{-\frac{1}{m}}$ )

$$\begin{aligned} \frac{S}{n} \left( c_n \left( \frac{\pi}{\tau} \right)^{2(m-1)} \sum_{l \leq N_n} l^{2(m-1)} - \left( \frac{\pi}{\tau} \right)^{m-2} \sum_{l \leq N_n} l^{m-2} \right) &= -c_n R, \\ \frac{S}{n} \left( c_n \left( \frac{\pi}{\tau} \right)^{2(m-1)} \frac{N_n^{2m-1}}{2m-1} - \left( \frac{\pi}{\tau} \right)^{m-2} \frac{N_n^{m-1}}{m-1} \right) &= -c_n R(1 + o(1)), \\ \frac{S}{n} \frac{\tau}{\pi} c_n^{-\frac{m-1}{m}} \left( \frac{1}{2m-1} - \frac{1}{m-1} \right) &= -c_n R(1 + o(1)). \end{aligned}$$

Finally, for the solution of (1.20), we can write

$$c_n^* = \alpha_n^*(1 + o(1)), \quad \alpha_n^* = \left[ \frac{S}{nR} \frac{\tau}{\pi} \frac{m}{(2m-1)(m-1)} \right]^{\frac{m}{2m-1}}. \quad (1.21)$$

Now, using the identity ( $\beta \in \mathcal{N}$ ,  $\beta > 1$ )

$$\sum_{l > N_n} \frac{1}{l^\beta} = \frac{1}{N_n^{\beta-1}} \int_1^{+\infty} \frac{1}{x^\beta} dx \cdot (1 + o(1)), \quad n \rightarrow +\infty,$$

for  $\beta = 2$

$$\sum_{l > N_n} \frac{1}{l^2} = \frac{1}{N_n} \cdot (1 + o(1)), \quad n \rightarrow +\infty,$$

calculate

$$\begin{aligned} H(c_n^*) &= \frac{S}{n} \left[ (c_n^*)^2 \left( \frac{\pi}{\tau} \right)^{2(m-1)} \frac{N_n^{2m-1}}{2m-1} - 2c_n^* \left( \frac{\pi}{\tau} \right)^{m-2} \frac{N_n^{m-1}}{m-1} - \left( \frac{\tau}{\pi} \right)^2 \frac{1}{N_n} \right] (1 + o(1)) \\ &+ (c_n^*)^2 R = \frac{S}{n} \frac{\tau}{\pi} \left[ (c_n^*)^2 \frac{(c_n^*)^{-\frac{2m-1}{m}}}{2m-1} - 2c_n^* \frac{(c_n^*)^{-\frac{m-1}{m}}}{m-1} - (c_n^*)^{\frac{1}{m}} \right] (1 + o(1)) + (c_n^*)^2 R = \\ &= \frac{S}{n} \frac{\tau}{\pi} (c_n^*)^{\frac{1}{m}} \frac{-2m^2}{(2m-1)(m-1)} (1 + o(1)) + (c_n^*)^2 R = \\ &= (-2m)R (c_n^*)^{\frac{1}{m}} (c_n^*)^{\frac{2m-1}{m}} (1 + o(1)) + (c_n^*)^2 R = \\ &= -(2m-1)(\alpha_n^*)^2 R(1 + o(1)), \end{aligned}$$

where we have used the relation (1.21). Now, choosing the sequence  $c_n = \alpha_n^*$  for the definition of the estimator in (1.18), we obtain from (1.19)

$$\begin{aligned} \sup_{\Lambda \in \mathcal{F}_m^{per}(R,S)} \left( \mathbf{E}_\Lambda \|\tilde{\Lambda}_n - \Lambda\|^2 - \mathbf{E}_\Lambda \|\hat{\Lambda}_n - \Lambda\|^2 \right) &\leq \\ &\leq -(2m-1)(\alpha_n^*)^2 R(1 + o(1)) + \frac{C}{n} \max_l \frac{|1 - |\tilde{K}_{2l,n}|^2|}{\left( \frac{2\pi l}{\tau} \right)^m}. \end{aligned} \quad (1.22)$$

If we show that

$$\frac{1}{n} \max_l \frac{|1 - \tilde{K}_{l,n}^2|}{\left(\frac{\pi l}{\tau}\right)^m} = o(n^{-\frac{2m}{2m-1}}), \quad (1.23)$$

then, since

$$\Pi = (2m - 1)(\alpha_n^*)^2 R n^{\frac{2m}{2m-1}},$$

we get from (1.22)

$$\overline{\lim}_{n \rightarrow +\infty} n^{\frac{2m}{2m-1}} \sup_{\Lambda \in \mathcal{F}_m^{per}(R,S)} \left( \mathbf{E}_\Lambda \|\tilde{\Lambda}_n - \Lambda\|^2 - \mathbf{E}_\Lambda \|\hat{\Lambda}_n - \Lambda\|^2 \right) \leq -\Pi.$$

This combined with the proposition will end the proof. To prove (1.23) recall that

$$\tilde{K}_{l,n} = \left( 1 - \left| \frac{\pi l}{\tau} \right|^m \alpha_n^* \right)_+, \quad \alpha_n^* = \left[ \frac{S \tau}{nR \pi} \frac{m}{(2m-1)(m-1)} \right]^{\frac{m}{2m-1}}.$$

Therefore, for  $m > 1$  we have

$$\frac{1}{n} \max_l \frac{|1 - \tilde{K}_{l,n}^2|}{\left(\frac{\pi l}{\tau}\right)^m} \leq \frac{2}{n} \max_l \frac{1 - \tilde{K}_{l,n}}{\left(\frac{\pi l}{\tau}\right)^m} = \frac{2}{n} \alpha_n^* = \frac{C}{n^{\frac{3m-1}{2m-1}}} = o(n^{-\frac{2m}{2m-1}}).$$

■

### 1.4.3 Proof of Proposition 1.3

Consider the following minimax risk

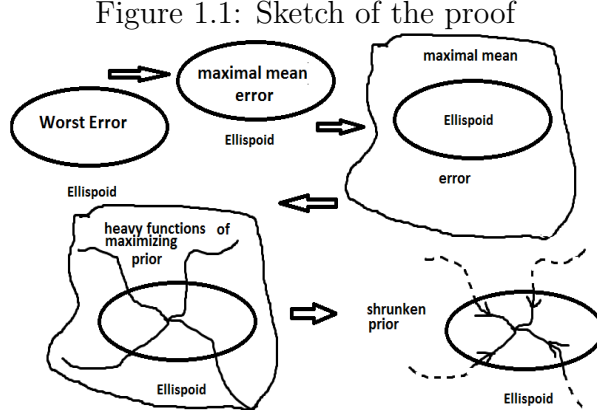
$$\mathcal{R}_n = \inf_{\bar{\Lambda}_n} \sup_{\mathcal{F}_m^{per}(R,S)} \left( \mathbf{E}_\Lambda \|\bar{\Lambda}_n - \Lambda\|^2 - \mathbf{E}_\Lambda \|\hat{\Lambda}_n - \Lambda\|^2 \right),$$

where the inf is taken over all possible estimators  $\bar{\Lambda}_n = \bar{\Lambda}_n(t, \mathbf{X}^n)$  (functions  $\bar{\Lambda}_n(t, \cdot)$  that are measurable w.r.t. the second variable). We have to prove that

$$\lim_{n \rightarrow +\infty} n^{\frac{2m}{2m-1}} \mathcal{R}_n \geq -\Pi.$$

The proof follows the main steps from [13]. The following graphic represents the sketch of the proof.

- A. Reduction of the minimax risk to a Bayes risk.
- B. Choosing a parametric family of functions, where the Bayesian prior distribution is concentrated (the heavy functions).



- C. Calculation of the Bayes risk for the least favorable prior.
- D. Shrinking the least favorable prior to fit the ellipsoid.

In the part C, in the problem of Bayes risk maximization, we replace the set of all prior distributions concentrated on the ellipsoid  $\mathcal{F}_m^{per}(R, S)$  by the set of all prior distributions concentrated on that ellipsoid in *mean*. Thus we enlarge the class of possible prior distributions, as it is represented in the third graphic of the Figure 1.1. Hence, the necessity of the part D (see Remark 1.4).

### Reduction to a Bayes risk of a parametric family

For each estimator  $\Lambda_n(t)$  there exists an estimator  $\bar{\Lambda}_n \in \mathcal{F}_m^{per}(R, S)$  so that (see also Section 3.3.2, [40])

$$\mathbf{E}_\Lambda \|\Lambda_n - \Lambda\|^2 \geq \mathbf{E}_\Lambda \|\bar{\Lambda}_n - \Lambda\|^2.$$

Indeed, if  $\Lambda_n \notin L_2[0, \tau]$  then this inequality is trivial for all  $\bar{\Lambda}_n \in \mathcal{F}_m^{per}(R, S)$  and if  $\Lambda_n \in L_2[0, \tau]$  then we can take as  $\bar{\Lambda}_n$  the projection of the estimator  $\Lambda_n$  on the closed, convex set  $\mathcal{F}_m^{per}(R, S)$ . Therefore

$$\mathcal{R}_n \geq \inf_{\bar{\Lambda}_n \in \mathcal{F}_m^{per}(R, S)} \sup_{\hat{\Lambda}_n \in \mathcal{F}_m^{per}(R, S)} \left( \mathbf{E}_\Lambda \|\bar{\Lambda}_n - \Lambda\|^2 - \mathbf{E}_\Lambda \|\hat{\Lambda}_n - \Lambda\|^2 \right).$$

For each parametric family of mean functions

$$\{\Lambda^\theta = \Lambda(\theta, t), t \in [0, \tau], \theta \in \Theta \subset \ell^2\},$$

denote

$$\Theta_0 = \{\theta \in \Theta, \Lambda^\theta \in \mathcal{F}_m^{per}(R, S)\}, \quad \Theta_0^c = \Theta \setminus \Theta_0,$$

and suppose that there is a probability measure  $\mathbf{Q}(d\theta)$  defined on the set  $\Theta$ . The subset  $\Theta$  of the Hilbert space  $\ell^2$ , as well as the measure  $\mathbf{Q}(\theta)$ , will be chosen latter. Hence

$$\begin{aligned}
\mathcal{R}_n &\geq \inf_{\bar{\Lambda}_n \in \mathcal{F}_m^{per}(R,S)} \sup_{\Lambda \in \mathcal{F}_m^{per}(R,S)} \left( \mathbf{E}_\Lambda \|\bar{\Lambda}_n - \Lambda\|^2 - \mathbf{E}_\Lambda \|\hat{\Lambda}_n - \Lambda\|^2 \right) \\
&\geq \inf_{\bar{\Lambda}_n \in \mathcal{F}_m^{per}(R,S)} \sup_{\Lambda^\theta \in \mathcal{F}_m^{per}(R,S)} \left( \mathbf{E}_\theta \|\bar{\Lambda}_n - \Lambda^\theta\|^2 - \mathbf{E}_\theta \|\hat{\Lambda}_n - \Lambda^\theta\|^2 \right) \\
&\geq \inf_{\bar{\Lambda}_n \in \mathcal{F}_m^{per}(R,S)} \int_{\Theta_0} \left( \mathbf{E}_\theta \|\bar{\Lambda}_n - \Lambda^\theta\|^2 - \mathbf{E}_\theta \|\hat{\Lambda}_n - \Lambda^\theta\|^2 \right) \mathbf{Q}(d\theta) \\
&= \inf_{\bar{\Lambda}_n \in \mathcal{F}_m^{per}(R,S)} \mathbb{E} \left( \|\bar{\Lambda}_n - \Lambda^\theta\|^2 - \|\hat{\Lambda}_n - \Lambda^\theta\|^2 \right) \\
&\quad - \sup_{\bar{\Lambda}_n \in \mathcal{F}_m^{per}(R,S)} \int_{\Theta_0^c} \left( \mathbf{E}_\theta \|\bar{\Lambda}_n - \Lambda^\theta\|^2 - \mathbf{E}_\theta \|\hat{\Lambda}_n - \Lambda^\theta\|^2 \right) \mathbf{Q}(d\theta) \\
&= \mathcal{R}_n^* - \mathcal{R}_n^0,
\end{aligned}$$

where  $\mathbb{E}$  is the expectation with respect to the measure  $\mathbf{Q}(d\theta)\mathbf{P}^n(dx)$  and the last equality is a notation.

### Bayes risk evaluation of an auxiliary parametric family

In this subsection we are interested in the evaluation of the term  $\mathcal{R}_n^*$ . Denote

$$r_n(\Lambda^\theta, \bar{\Lambda}_n) = \mathbb{E} \left( \|\bar{\Lambda}_n - \Lambda^\theta\|^2 - \|\hat{\Lambda}_n - \Lambda^\theta\|^2 \right), \text{ then, } \mathcal{R}_n^* = \inf_{\bar{\Lambda}_n \in \mathcal{F}_m^{per}(R,S)} r_n(\Lambda^\theta, \bar{\Lambda}_n).$$

Using the Parseval's equality for the basis (1.13), we can write

$$r_n(\Lambda^\theta, \bar{\Lambda}_n) = \sum_{k=0}^{+\infty} \left\{ \mathbb{E} |\bar{\Lambda}_{k,n} - \Lambda_k^\theta|^2 - \mathbb{E} |\hat{\Lambda}_{k,n} - \Lambda_k^\theta|^2 \right\},$$

with

$$\bar{\Lambda}_{k,n} = \int_0^\tau \bar{\Lambda}_n(t) \phi_k(t) dt, \quad \Lambda_k^\theta = \int_0^\tau \Lambda(\theta, t) \phi_l(t) dt, \quad \hat{\Lambda}_{k,n} = \int_0^\tau \hat{\Lambda}_n(t) \phi_k(t) dt.$$

The paper [1] suggested that the lower bound for this term can be obtained using the van Trees inequality. For various types of this inequality see [10]. The following one is called the  $L_2$ -type van Trees inequality ([10]) (for a special case of this inequality see [13], for the proof in the one dimensional parameter case see also [40]).



Suppose  $(\mathcal{X}, \mathcal{A}, \{\mathbf{P}_\theta, \theta \in \Theta\}, \nu)$  is a measurable space, where the probability measures are absolutely continuous w.r.t. the measure  $\nu(dx)$

$$\mathbf{P}_\theta \ll \nu, \theta \in \Theta, \quad f(x, \theta) = \frac{d\mathbf{P}_\theta}{d\nu}(x).$$

Here  $\Theta = \prod_{i=1}^{+\infty} [\alpha_i, \beta_i] \subset \ell^2$  is a hyper-rectangle. Introduce the following conditions

**I.**  $f(x, \theta)$  is measurable in  $(x, \theta)$  and for  $\nu$  almost all  $x \in \mathcal{X}$  the functions  $\theta_k \mapsto f(x, \theta)$  are absolutely continuous on the segment  $[\alpha_k, \beta_k]$ ,

**II.** The Fisher information is finite and integrable

$$I_k(\theta) = \int_{\mathcal{X}} \left( \frac{\frac{\partial f(x, \theta)}{\partial \theta_k}}{f(x, \theta)} \right)^2 f(x, \theta) \nu(dx) < +\infty, \quad \int_{\Theta} I_k(\theta) d\theta < +\infty, \quad k \in \mathcal{N},$$

**III.** The components of the vector  $\theta = (\theta_1, \theta_2, \dots)$  are independent random variables w.r.t. the Lebesgue measure, with the absolutely continuous densities  $p_k(x)$ ,  $x \in [\alpha_k, \beta_k]$ , so that  $p_k(\alpha_k) = p_k(\beta_k) = 0$  and

$$I_k = \int_{\alpha_k}^{\beta_k} \frac{[p'_k(x)]^2}{p_k(x)} dx < +\infty, \quad k \in \mathcal{N}.$$

Denote by  $\mathbf{Q}$  the distribution of the vector  $\theta$ .

**IV.** The function  $\psi(\theta)$  is such that  $\theta_k \mapsto \psi(\theta)$  are absolutely continuous on  $[\alpha_k, \beta_k]$ ,  $k \in \mathcal{N}$ , and  $\mathbf{E}_{\mathbf{Q}} \psi(\theta) < +\infty$ .

Now, suppose that the components of  $\mathbf{X}^n = (X_1, X_2, \dots, X_n)$  are independent, identically distributed according to the law  $\mathbf{P}_\theta$ . Denote the distribution of  $\mathbf{X}^n$  by  $\mathbf{P}^n(dx)$  and by  $\mathbb{E}$  the expectation w.r.t. the measure  $\mathbf{Q}(d\theta)\mathbf{P}^n(dx)$ .

**Theorem 1.2** (The van Trees inequality, [10]). *Under the assumptions I-IV, for all estimators  $\psi_n = \psi_n(\mathbf{X}^n)$  the following inequality holds*

$$\mathbb{E} |\psi_n - \psi(\theta)|^2 \geq \frac{\left[ \mathbf{E}_{\mathbf{Q}} \frac{\partial \psi(\theta)}{\partial \theta_k} \right]^2}{n \mathbf{E}_{\mathbf{Q}} I_k(\theta) + I_k}, \quad k \in \mathcal{Z}_+.$$

The proof of this inequality is given in the Appendix.

We use the van Trees inequality for the Poisson measures, to obtain an inequality for the Fourier coefficients of the mean function. Recall the form of the Fisher

information for a family of Poisson measures (1.1). Taking  $\psi(\theta) = \Lambda_k^\theta$ , and applying the van Trees inequality for the Poisson measures, we obtain the following inequality

$$\mathbb{E}|\bar{\Lambda}_{k,n} - \Lambda_k^\theta|^2 \geq \frac{\left[\mathbf{E}_{\mathbf{Q}} \frac{\partial \Lambda_k^\theta}{\partial \theta_k}\right]^2}{n\mathbf{E}_{\mathbf{Q}}\mathbf{I}_k(\theta) + \mathbf{I}_k}, \quad k \in \mathcal{Z}_+. \quad (1.24)$$

**Remark 1.3.** *We are going to choose a parametric family in a way to maximize the right-hand side of the above inequality. Such functions we call the heavy functions. The family of these functions will be the least favorable.*  $\diamond$

Consider the second order mean integrated Bayes risk for the a parametric family

$$\begin{aligned} r_n(\Lambda^\theta, \bar{\Lambda}_n) &= \sum_{k=0}^{+\infty} \left\{ \mathbb{E}|\bar{\Lambda}_{k,n} - \Lambda_k^\theta|^2 - \mathbb{E}|\hat{\Lambda}_{k,n} - \Lambda_k^\theta|^2 \right\} = \\ &= \sum_{k=0}^{+\infty} \left\{ \mathbb{E}|\bar{\Lambda}_{k,n} - \Lambda_k^\theta|^2 - \sigma_{k,n}^2 \right\}. \end{aligned} \quad (1.25)$$

To compute the variance  $\sigma_{k,n}^2 = \mathbb{E}|\hat{\Lambda}_{k,n} - \Lambda_k^\theta|^2$  introduce the notations

$$\pi_j(t) = X_j(t) - \Lambda(t), \quad g_k(t) = \int_t^\tau \phi_k(s) ds.$$

In the sequel, we are going to use the following property of stochastic integral

$$\mathbf{E}_\Lambda \left[ \int_0^\tau f(t) d\pi_j(t) \int_0^\tau g(t) d\pi_j(t) \right] = \int_0^\tau f(t)g(t) d\Lambda(t), \quad f, g \in L_2[0, \tau].$$

Further, in view of the integration by parts, we have

$$\hat{\Lambda}_{k,n} - \Lambda_k = \frac{1}{n} \sum_{j=1}^n \int_0^\tau \pi_j(t) \phi_k(t) dt = \frac{1}{n} \sum_{j=1}^n \int_0^\tau \left( \int_t^\tau \phi_k(s) ds \right) d\pi_j(t).$$

which implies that

$$\sigma_{k,n}^2 = \frac{1}{n} \mathbf{E}_{\mathbf{Q}} \int_0^\tau \left( \int_t^\tau \phi_k(s) ds \right)^2 d\Lambda(\theta, t) = \frac{1}{n} \mathbf{E}_{\mathbf{Q}} \int_0^\tau g_k^2(s) \lambda(\theta, t) dt. \quad (1.26)$$

The van Trees inequality (1.24) gives us the following lower bound

$$\begin{aligned} r_n(\Lambda_\theta, \bar{\Lambda}_n) &\geq \sum_{k=0}^{+\infty} \left( \frac{\left[\mathbf{E}_{\mathbf{Q}} \frac{\partial \Lambda_k^\theta}{\partial \theta_k}\right]^2}{n\mathbf{E}_{\mathbf{Q}}\mathbf{I}_k(\theta) + \mathbf{I}_k} - \sigma_{k,n}^2 \right) \\ &= \sum_{k=0}^{+\infty} \frac{\left[\mathbf{E}_{\mathbf{Q}} \frac{\partial \Lambda_k^\theta}{\partial \theta_k}\right]^2 - n\mathbf{E}_{\mathbf{Q}}\mathbf{I}_k(\theta)\sigma_{k,n}^2}{n\mathbf{E}_{\mathbf{Q}}\mathbf{I}_k(\theta) + \mathbf{I}_k} - \sum_{k=0}^{+\infty} \frac{\sigma_{k,n}^2 \mathbf{I}_k}{n\mathbf{E}_{\mathbf{Q}}\mathbf{I}_k(\theta) + \mathbf{I}_k}. \end{aligned} \quad (1.27)$$

The Fourier coefficients of the function from the parametric family can be written as

$$\Lambda_k^\theta = \int_0^\tau \Lambda(\theta, t) \phi_k(t) dt = \int_0^\tau \lambda(\theta, t) g_k(t) dt,$$

$$\frac{\partial \Lambda_k^\theta}{\partial \theta_k} = \int_0^\tau g_k(t) \frac{\partial \lambda(\theta, t)}{\partial \theta_k} dt = \int_0^\tau g_k(t) \frac{\partial(\ln \lambda(\theta, t))}{\partial \theta_k} \lambda(\theta, t) dt,$$

and the application of the the Cauchy-Schwarz inequality entails

$$\left[ \mathbf{E}_{\mathbf{Q}} \frac{\partial \Lambda_k^\theta}{\partial \theta_k} \right]^2 \leq \mathbf{E}_{\mathbf{Q}} \left[ \int_0^\tau g_k^2(t) \lambda(\theta, t) dt \right] \mathbf{E}_{\mathbf{Q}} I_k(\theta).$$

The equality in the Cauchy-Schwarz inequality possible if and only if, for some constant  $c \in \mathbf{R}$

$$\frac{\partial(\ln \lambda(\theta, t))}{\partial \theta_k} = c g_k(t), \quad k \in \mathcal{Z}_+.$$

Hence, the solutions of these differential equations with the terminal condition  $\Lambda(\tau) = S$  is

$$\lambda(\theta, t) = \frac{S}{\tau} \exp \left\{ \sum_{k=0}^{+\infty} \theta_k g_k(t) \right\}. \quad (1.28)$$

are *the heavy functions*, so the right-hand side of (1.24) will be maximal and for this parametric family

$$\int_0^\tau g_k^2(t) \lambda(\theta, t) dt = I_k(\theta), \quad (1.29)$$

therefore (1.27) becomes

$$\begin{aligned} r_n(\Lambda_\theta, \bar{\Lambda}_n) &\geq -\frac{1}{n} \sum_{k=1}^{+\infty} \frac{\mathbf{E}_{\mathbf{Q}} I_k(\theta) I_k}{n \mathbf{E}_{\mathbf{Q}} I_k(\theta) + I_k} - \frac{1}{n} \frac{\mathbf{E}_{\mathbf{Q}} I_0(\theta) I_0}{n \mathbf{E}_{\mathbf{Q}} I_0(\theta) + I_0} \geq \\ &\geq -\frac{1}{n} \sum_{k=1}^{+\infty} \frac{\mathbf{E}_{\mathbf{Q}} I_k(\theta) I_k}{n \mathbf{E}_{\mathbf{Q}} I_k(\theta) + I_k} - \frac{I_0}{n^2}. \end{aligned} \quad (1.30)$$

We have to impose conditions on the parametric set  $\Theta$  for the heavy functions  $\lambda(\theta, t)$  to fit the ellipsoid  $\mathcal{F}_m^{per}(R, S)$ . Consider the following set

$$\Theta_1 = \left\{ \theta : \left( \frac{S}{\tau} \right)^2 \sum_{k=1}^{+\infty} \left( \frac{\pi k}{\tau} \right)^{2(m-2)} \theta_k^2 \leq R \right\}. \quad (1.31)$$

We are going to show that asymptotically, if  $\ell_2$  norm of the parameter  $\theta$  tends to zero over the set  $\Theta_1$ , then the parametric family asymptotically fits the ellipsoid  $\mathcal{F}_m^{per}(R, S)$ . We need also to calculate the asymptotics of the Fisher information of the mentioned parametric family. The following lemma is similar to Lemma 9 from [13].

**Lemma 1.2.** *Consider the parametric family (1.28) with  $\theta \in \Theta_1 \cap \Theta_n$ , where  $\Theta_1$  is defined in (1.31) and*

$$\Theta_n = \left\{ \theta : \sum_{k=0}^{+\infty} \theta_k^2 \leq \varepsilon_n \right\}, \quad \varepsilon_n = o((\ln n)^{-1}).$$

Then, as  $n \rightarrow +\infty$ , uniformly in  $\theta \in \Theta_1 \cap \Theta_n$  the following properties hold

1.  $\Lambda(\theta, \tau) = S + o(1)$ ,
2.  $I_k(\theta) \rightarrow \frac{S}{\tau} \left(\frac{\tau}{\pi k}\right)^2$ , for all  $k \in \mathcal{N}$ ,
3.  $\int_0^\tau [\lambda^{(m-1)}(\theta, t)]^2 dt \leq R + o(1)$ .

*Proof.*

1. Using the Cauchy-Schwarz inequality we can write

$$\left| \sum_{k=0}^{+\infty} g_k(t) \theta_k \right| \leq C \sum_{k=0}^{+\infty} \left| \frac{\theta_k}{1+k} \right| \leq C \|\theta\|_{\ell_2} \leq C \varepsilon_n \rightarrow 0,$$

hence from (1.28) we obtain the following uniform convergence over  $\Theta_1 \cap \Theta_n$  as  $n \rightarrow +\infty$

$$\lambda(\theta, t) \rightarrow \frac{S}{\tau},$$

which implies

$$\Lambda(\theta, t) \rightarrow S.$$

2. According to (1.29)

$$I_k(\theta) = \int_0^\tau g_k^2(t) \lambda(\theta, t) dt \rightarrow \frac{S}{\tau} \int_0^\tau g_k^2(t) dt = \frac{S}{\tau} \left(\frac{\tau}{\pi k}\right)^2.$$

3. Following a suggestion from [13] (Lemma 9), we are going to use a well known fact from [39]. For a function  $g(\cdot)$  periodic on  $[0, \tau]$  the following inequality is true for  $0 \leq s < k$

$$\sup_{t \in [0, \tau]} |g^{(s)}(t)|^2 \leq C \|g^{(k)}\|^{2p} \|g\|^{2(1-p)}, \quad p = \frac{2s+1}{2k}. \quad (1.32)$$

Denote  $f(t) = \sum_{k=0}^{+\infty} \theta_k g_k(t)$ ,  $F(t) = e^{f(t)}$ . For the  $(m-1)$ th derivative of  $F(t)$  we have the following representation

$$F^{(m-1)}(t) = F(t)[f^{(m-1)}(t) + P(f'(t), \dots, f^{(m-2)}(t))], \quad (1.33)$$

where  $P(\cdot)$  is a polynomial. Then, using the Parseval's equality, we find (see (1.31))

$$\left\| \frac{S}{\tau} f^{(m-1)} \right\|^2 = \left( \frac{S}{\tau} \right)^2 \sum_{k=1}^{+\infty} \left( \frac{\pi k}{\tau} \right)^{2(m-2)} \theta_k^2 \leq R, \quad \|f'\|^2 = \sum_{k=0}^{+\infty} \theta_k^2 \leq \varepsilon_n \rightarrow 0.$$

Applying (1.32) for  $g = f'$  with  $k = m-1$  and  $1 \leq s \leq m-2$  gives us

$$\sup_{t \in [0, \tau]} |f^{(s)}(t)|^2 \leq C \|f^{(m-1)}\|^{\frac{2m-3}{m-1}} \|f'\|^{\frac{1}{m-1}} \rightarrow 0.$$

Since, according to the first statement of this Lemma,

$$F(t) \rightarrow 1$$

uniformly, as  $n \rightarrow +\infty$ , then, combining previous statements with (1.33), we get

$$\int_0^\tau [\lambda^{(m-1)}(\theta, t)]^2 = \|\lambda^{(m-1)}\|^2 \leq R + o(1).$$

■

Using the Lemma 1.2, we can transform (1.30) to

$$r_n(\Lambda_\theta, \bar{\Lambda}_n) \geq -\frac{1}{n} \sum_{k=1}^{+\infty} \frac{\frac{S}{\tau} \left( \frac{\tau}{\pi k} \right)^2 \mathbf{I}_k}{n \frac{S}{\tau} \left( \frac{\tau}{\pi k} \right)^2 + \mathbf{I}_k} (1 + o(1)) - \frac{\mathbf{I}_0}{n^2}. \quad (1.34)$$

### Least favorable Bayesian risk calculation

In this subsection we are going to carry out C and D parts of the proof, namely, to choose the least favorable prior distribution  $\mathbf{Q}(\theta)$  and to calculate the corresponding Bayesian risk. For a general overview of the Bayes risk maximization problem we refer to [17].

Recall that  $\mathbf{Q}(\theta)$  is the distribution of the vector  $\theta$  and the conditions on  $\mathbf{Q}(\theta)$  are that it has to be concentrated on some hyper-rectangle (Theorem 1.2) and the vector  $\theta$  has to satisfy the inequality (Lemma 1.2)

$$\left( \frac{S}{\tau} \right)^2 \sum_{k=1}^{+\infty} \left( \frac{\pi k}{\tau} \right)^{2(m-2)} \theta_k^2 \leq R.$$

We have to choose a prior distribution satisfying these two conditions, which maximizes the functional (see (1.34))

$$-\frac{1}{n} \frac{S}{\tau} \sum_{k=1}^{+\infty} \frac{1}{n \frac{S}{\tau} \frac{1}{I_k} + \left(\frac{\pi k}{\tau}\right)^2},$$

where  $I_k$  is the Fisher information of the co-ordinate  $\theta_k$ .

First, independence is less favorable ([17]), that is, the co-ordinates  $\theta_k$  must be chosen independently  $\mathbf{Q}(d\theta) = \prod_{j=1}^{+\infty} p_j(\theta_j) d\theta_j$ . Then, for each random variable  $\xi$  with the density  $p(x)$ , the variance  $\sigma^2$  and the Fisher information  $I = \int_R \frac{[p'(x)]^2}{p(x)} dx$ , the following inequality  $\sigma^2 \cdot I \geq 1$  holds, with the equality if and only if  $\xi$  is normally distributed. Therefore, denoting  $\varphi_k = \frac{1}{I_k}$ , we have to choose  $\theta_k$  distributed  $\mathcal{N}(0, \varphi_k)$ . Thus, we derive the maximization problem of the functional

$$S(\varphi) = -\frac{1}{n} \frac{S}{\tau} \sum_{k=1}^{+\infty} \frac{1}{n \frac{S}{\tau} \varphi_k + \left(\frac{\pi k}{\tau}\right)^2},$$

over the set

$$\left\{ \varphi = (\varphi_k)_{k \geq 1} : \left(\frac{S}{\tau}\right)^2 \sum_{k=1}^{+\infty} \left(\frac{\pi k}{\tau}\right)^{2(m-2)} \varphi_k \leq R \right\}.$$

Applying the Lagrange multipliers method, first, for a constant  $\mu^2$ ,  $\mu > 0$  we have to solve the equality

$$\frac{1}{n} \frac{S}{\tau} \frac{n \frac{S}{\tau}}{\left[n \frac{S}{\tau} \varphi_k + \left(\frac{\pi k}{\tau}\right)^2\right]^2} = \mu^2 \left(\frac{S}{\tau}\right)^2 \left(\frac{\pi k}{\tau}\right)^{2(m-2)},$$

whence

$$\varphi_k^0 = \frac{1}{n} \frac{\tau}{S} \left(\frac{\pi k}{\tau}\right)^2 \left(\frac{1}{\mu} \left(\frac{\tau}{\pi k}\right)^m - 1\right)_+,$$

and denoting  $W = \left(\frac{1}{\mu}\right)^{\frac{1}{m}} \frac{\tau}{\pi}$ , we obtain the solution of the above mentioned equation

$$\varphi_k^0 = \frac{1}{n} \frac{\tau}{S} \left(\frac{\pi k}{\tau}\right)^2 \left(\left(\frac{W}{k}\right)^m - 1\right)_+, \quad k \in \mathcal{N}. \quad (1.35)$$

The number  $W$  can be found from the equation

$$\left(\frac{S}{\tau}\right)^2 \sum_{k=1}^{+\infty} \left(\frac{\pi k}{\tau}\right)^{2(m-2)} \varphi_k^0 = R.$$

Actually, since  $\varphi_k^0$  depends on  $n$ , the number  $W = W_n$  also depends on  $n$ , hence, as  $n \rightarrow +\infty$ , we can find the asymptotics of  $W_n$ .

$$\begin{aligned} \frac{1}{n} \frac{S}{\tau} \sum_{k=1}^{+\infty} \left(\frac{\pi k}{\tau}\right)^{2(m-1)} \left(\left(\frac{W}{k}\right)^m - 1\right)_+ &= R, \\ \frac{1}{n} \frac{S}{\tau} \sum_{k \leq W} \left(\frac{\pi k}{\tau}\right)^{2(m-1)} \left(\left(\frac{W}{k}\right)^m - 1\right) &= R, \\ \frac{1}{n} \frac{S}{\tau} \left(\frac{\pi}{\tau}\right)^{2(m-1)} \sum_{k \leq W} [W^m k^{m-2} - k^{2(m-1)}] &= R, \\ \frac{1}{n} \frac{S}{\tau} \left(\frac{\pi}{\tau}\right)^{2(m-1)} \left[ W^m \frac{W^{m-1}}{m-1} - \frac{W^{2m-1}}{2m-1} \right] (1 + o(1)) &= R. \end{aligned}$$

Finally for  $W = W_n$  we have the following asymptotic behavior, as  $n \rightarrow +\infty$ ,

$$W_n = \frac{\tau}{\pi} \left[ \frac{n\pi R (m-1)(2m-1)}{S m} \right]^{\frac{1}{2m-1}} (1 + o(1)). \quad (1.36)$$

Now, we are going to calculate the maximal value of the functional  $S(\varphi)$ . Using

(1.35) we calculate

$$\begin{aligned}
S(\varphi_0) &= -\frac{1}{n} \frac{S}{\tau} \sum_{k=1}^{+\infty} \frac{1}{n \frac{S}{\tau} \varphi_k^0 + \left(\frac{\pi k}{\tau}\right)^2} = \\
&= -\frac{1}{n} \frac{S}{\tau} \left[ \sum_{k \leq W} \frac{1}{\left(\frac{\pi k}{\tau}\right)^2 \left(\frac{W}{k}\right)^m} + \sum_{k > W} \left(\frac{\tau}{\pi k}\right)^2 \right] = \\
&= -\frac{1}{n} \frac{S}{\tau} \left(\frac{\tau}{\pi}\right)^2 \left[ \frac{1}{W^m} \sum_{k \leq W} k^{m-2} + \sum_{k > W} \frac{1}{k^2} \right] (1 + o(1)) = \\
&= -\frac{1}{n} \frac{S}{\tau} \left(\frac{\tau}{\pi}\right)^2 \left[ \frac{1}{W^m} \sum_{k \leq W} \frac{W^{m-1}}{m-1} + \frac{1}{W} \right] (1 + o(1)) = \\
&= -\frac{1}{n} \frac{S}{\tau} \left(\frac{\tau}{\pi}\right)^2 \frac{1}{W} \frac{m}{m-1} (1 + o(1)) = \\
&= -R(2m-1) \frac{S}{n\pi R} \frac{m}{(m-1)(2m-1)} \frac{\tau}{\pi W} (1 + o(1)) = \\
&= -R(2m-1) n^{-\frac{2m}{2m-1}} \left[ \frac{S}{\pi R} \frac{m}{(m-1)(2m-1)} \right]^{\frac{2m}{2m-1}} (1 + o(1)) = \\
&= -n^{-\frac{2m}{2m-1}} \Pi(1 + o(1)).
\end{aligned}$$

Where we have used (1.36) and the definition of  $\Pi$ . Lemma 1.2 combining with (1.30) gives us

$$r_n(\Lambda^\theta, \bar{\Lambda}_n) \geq -n^{-\frac{2m}{2m-1}} \Pi(1 + o(1)),$$

for all estimators  $\bar{\Lambda}_n$ , hence

$$\mathcal{R}_n^* \geq -n^{-\frac{2m}{2m-1}} \Pi(1 + o(1)).$$

If we show that  $\mathcal{R}_n^0 = o(n^{-\frac{2m}{2m-1}})$ , then we can write

$$\mathcal{R}_n \geq -n^{-\frac{2m}{2m-1}} \Pi(1 + o(1)),$$

which was the statement of the Proposition 1.3.

### Completion of the proof of the Proposition 1.3

The finish the proof of the Proposition 1.3 we have to show that for the residual term we have  $\mathcal{R}_n^0 = o(n^{-\frac{2m}{2m-1}})$ . We use a technique from [40] (Section 3.3.2).



Denote  $d = \sup_{\Lambda \in \mathcal{F}_m^{per}(R,S)} \|\Lambda\|$ . Then, for  $\Theta_0 = \{\theta : \Lambda(\theta, \cdot) \in \mathcal{F}_m^{per}(R,S)\}$  (see (1.31))

$$\begin{aligned} |\mathcal{R}_n^0| &\leq \sup_{\bar{\Lambda}_n \in \mathcal{F}_m^{per}(R,S)} \int_{\Theta_0^c} \left| \mathbf{E}_\Lambda \int_0^\tau (\bar{\Lambda}_n(t) - \Lambda(\theta, t))^2 dt - \frac{1}{n} \int_0^\tau \Lambda(\theta, t) dt \right| \mathbf{Q}(d\theta) \\ &\leq 2 \int_{\Theta_0^c} (d^2 + \|\Lambda_\theta\|^2) \mathbf{Q}(d\theta) + \frac{\sqrt{\tau}}{n} \int_{\Theta_0^c} \|\Lambda_\theta\|^2 \mathbf{Q}(d\theta). \end{aligned}$$

By the Cauchy-Schwarz inequality we have

$$\begin{aligned} \int_{\Theta_0^c} \|\Lambda_\theta\|^2 \mathbf{Q}(d\theta) &\leq \left( \int_{\Theta_0^c} \|\Lambda_\theta\|^4 \mathbf{Q}(d\theta) \right)^{\frac{1}{2}} \left( \int_{\Theta_0^c} \mathbf{Q}(d\theta) \right)^{\frac{1}{2}}, \\ \int_{\Theta_0^c} \|\Lambda_\theta\|^2 \mathbf{Q}(d\theta) &\leq (\mathbf{E}_\mathbf{Q} \|\Lambda_\theta\|^4)^{\frac{1}{2}} \sqrt{\mathbf{Q}(\Theta_0^c)}. \end{aligned}$$

Finally, for the residual term we obtain

$$|\mathcal{R}_n^0| \leq 2 \left[ d^2 \mathbf{Q}(\Theta_0^c) + (\mathbf{E}_\mathbf{Q} \|\Lambda_\theta\|^4)^{\frac{1}{2}} \sqrt{\mathbf{Q}(\Theta_0^c)} \right] + \frac{\sqrt{\tau}}{n} (\mathbf{E}_\mathbf{Q} \|\Lambda_\theta\|^4)^{\frac{1}{2}} \sqrt{\mathbf{Q}(\Theta_0^c)}.$$

Since  $d^2$  and  $\mathbf{E}_\mathbf{Q} \|\Lambda_\theta\|^4$  are bounded then to prove  $\mathcal{R}_n^0 = o(n^{-\frac{2m}{2m-1}})$  we have to prove that

$$\mathbf{Q}(\Theta_0^c) = o(n^{-\frac{4m}{2m-1}}).$$

**Remark 1.4.** We have mentioned earlier that in the Bayes risk maximization problem we looked for the least favorable prior distribution not among the probability distributions concentrated on the ellipsoid  $\mathcal{F}_m^{per}(R,S)$ , but among the probability distributions concentrated on that ellipsoid in mean. Hence, the obtained normal distributions for  $\theta_k$  satisfy the condition

$$\sum_{k=1}^{+\infty} \left( \frac{\pi k}{\tau} \right)^{2(m-2)} \theta_k^2 \leq R$$

only in mean. To shrink the least favorable prior to fit the ellipsoid we have to slightly modify the variances of the components  $\theta_k$ . We will consider the  $\theta_k$  normally distributed with the zero mean and the variance  $\varphi_k^\delta$ ,  $0 < \delta < 1$ , which are approximations of the variance (1.35).  $\diamond$

**Remark 1.5.** In the Theorem 1.2 we need  $\theta_k$  to be concentrated on bounded intervals, which is not the case if we take these random variables normally distributed. We are going to take bounded random variables, which asymptotically has the same properties as the chosen least favorable normal distributions.  $\diamond$

Consider i.i.d. random variables with zero mean, unit variance and density  $p(x)$ , such that for some  $G > 0$

$$|\xi_k| < G \text{ and } I = \int_R \frac{[p'(x)]^2}{p(x)} dx = 1 + \delta, \quad 0 < \delta < 1.$$

Suppose that the measure  $\mathbf{Q}(d\theta)$  is induced by the random variables  $\theta_k = \xi_k \sqrt{\varphi_k^\delta}$ , where

$$\varphi_k^\delta = \frac{1}{n} \frac{\tau}{S} \left( \frac{\pi k}{\tau} \right)^2 \left( \left( \frac{W(1-\delta)}{k} \right)^m - 1 \right)_+.$$

In this case the main part of the lower bound (1.34) will not change

$$S(\varphi_k^\delta) = \frac{1}{1-\delta} S(\varphi_k^0)(1 + o(1))$$

and since

$$\left( \frac{S}{\tau} \right)^2 \sum_{k=1}^{+\infty} \left( \frac{\pi k}{\tau} \right)^{2(m-2)} \varphi_k^0 = R$$

we find

$$\left( \frac{S}{\tau} \right)^2 \sum_{k=1}^{+\infty} \left( \frac{\pi k}{\tau} \right)^{2(m-2)} \mathbf{E} \mathbf{Q} \theta_k^2 = (1-\delta)R(1 + o(1)).$$

Using this equality calculate (see (1.31))

$$\begin{aligned} \mathbf{Q}(\Theta_1^c) &= \mathbf{Q} \left\{ \left( \frac{S}{\tau} \right)^2 \sum_{k=1}^{+\infty} \left( \frac{\pi k}{\tau} \right)^{2(m-2)} \theta_k^2 > R \right\} = \\ &= \mathbf{Q} \left\{ \left( \frac{S}{\tau} \right)^2 \sum_{k=1}^{+\infty} \left( \frac{\pi k}{\tau} \right)^{2(m-2)} (\theta_k^2 - \mathbf{E} \theta_k^2) > \delta R \right\} \leq e^{-\frac{2\delta^2 R^2}{P}}, \end{aligned}$$

here, the last part is due to the Hoeffding's inequality (see, for example, Lemma 2.8, [40]) and

$$\begin{aligned} P &= G^4 \left( \frac{S}{\tau} \right)^2 \sum_{k \leq W} \left( \frac{\pi k}{\tau} \right)^{4(m-2)} (\varphi_k^\delta)^2 = \\ &= \frac{G^2}{n^2} \sum_{k \leq W} \left( \frac{\pi k}{\tau} \right)^{4(m-1)} \left[ \frac{W^m (1-\delta)^m}{k^m} - 1 \right]^2 \leq C \frac{W^{4m-3}}{n^2}, \end{aligned}$$

hence  $P = O(n^{-\frac{1}{2m-1}})$ , which in turn gives us  $\mathbf{Q}(\Theta_1^c) \leq \exp(-Cn^{\frac{1}{2m-1}})$ .

For sufficiently large  $n$  we have  $\mathbf{Q}(\Theta_n) = 1$ , according to the choice of  $\varphi_k^\delta$  and  $\varepsilon_n$  (see Lemma 1.2). Hence  $\mathbf{Q}(\Theta_n \cap \Theta_1) = \mathbf{Q}(\Theta_1)$ . Using the statement 3 from Lemma 1.2, we get

$$\lim_{n \rightarrow +\infty} \mathbf{Q}(\Theta_0) = \lim_{n \rightarrow +\infty} \mathbf{Q}(\Theta_n \cap \Theta_1) = \lim_{n \rightarrow +\infty} \mathbf{Q}(\Theta_1),$$

therefore

$$\mathbf{Q}(\Theta_0^c) \leq \exp(-Cn^{\frac{1}{2m-1}}).$$

To obtain the statement of the Proposition 1.3 let  $n \rightarrow +\infty$  and  $\delta \rightarrow 0$  in the following inequality

$$\mathcal{R}_n \geq -\frac{n^{-\frac{2m}{2m-1}}}{1-\delta} \Pi(1+o(1)) + o(n^{-\frac{2m}{2m-1}}).$$

■

# Chapter 2

## Approximation of the solution of a BSDE

### 2.1 Stochastic differential equations

To introduce stochastic differential equations we need the definition of the Wiener process. Consider a probability space  $(\Omega, \mathcal{F}, \mathbf{P})$ .

**Definition 2.1** ([29]). *A stochastic process  $(W_t)$ ,  $0 \leq t \leq T$  is called Wiener process if*

1.  $W_0 = 0$  almost sure,
2. It has independent increments,
3.  $W_t - W_s$  has normal  $\mathcal{N}(0, |t - s|)$  distribution,
4. For almost all  $\omega \in \Omega$  the functions  $W_t = W_t(\omega)$  are continuous on  $0 \leq t \leq T$ .

Denote the natural filtration of the Wiener process  $\mathcal{F}_t^W = \sigma(W_s, s \leq t)$ ,  $0 \leq t \leq T$ . The Wiener process is a continuous  $(W_t, \mathcal{F}_t^W)$  martingale

$$\mathbf{E}(W_t | \mathcal{F}_s^W) = W_s, \quad t > s,$$

moreover

$$\mathbf{E}((W_t - W_s)^2 | \mathcal{F}_s^W) = t - s.$$

Like any process with independent increments the Wiener process is a Markov process

$$\mathbf{E} [f(W_{t+s}) | \mathcal{F}_s^W] = \mathbf{E} [f(W_{t+s}) | W_s],$$

for any measurable function with  $\sup_x |f(x)| < +\infty$ .

Consider a non-homogeneous stochastic differential equation (SDE)

$$dX_t = S(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = x_0, \quad t \in [0, T]. \quad (2.1)$$

**Proposition 2.1** ([29]). *Under the conditions  $(\forall x, y \in \mathbf{R}, \forall t \in [0, T])$*

1.  $|S(t, x) - S(t, y)|^2 + |\sigma(t, x) - \sigma(t, y)|^2 \leq L|x - y|^2$ , (*Lipschitz condition*)
2.  $S^2(t, x) + \sigma^2(t, x) \leq L(1 + x^2)$ , (*no faster than linear growth condition*)

*the stochastic differential equation (2.1) has a unique, strong solution, that is, there exists a  $(\mathcal{F}_t^W)$ -adapted stochastic process  $\{X_t, t \in [0, T]\}$  on the probability space  $(\Omega, \mathcal{F}, \mathbf{P})$ , which satisfies with probability 1*

$$X_t = x_0 + \int_0^t S(u, X_u)du + \int_0^t \sigma(u, X_u)dW_u, \quad t \in [0, T].$$

*Uniqueness means that for another process  $\{Y_t, t \in [0, T]\}$  with such properties one has*

$$\mathbf{P}\left(\sup_{0 \leq t \leq T} |X_t - Y_t| > 0\right) = 0.$$

**Remark 2.1.** *Since  $X_0 = x_0$  is constant then it has all moments*

$$\mathbf{E}|X_t|^m < +\infty, \quad m \in \mathcal{N}.$$

◇

The theory of backward stochastic differential equations relies on the martingale representation theorem.

## 2.2 Backward stochastic differential equations

In this section we are going to introduce the notion of the backward stochastic differential equation (BSDE) in the Markovian case. In this case the backward stochastic differential equations are called forward backward stochastic differential equations (FBSDE).

### 2.2.1 Markovian case

Let us recall what is the BSDE in the Markovian case. Suppose that we are given a filtered probability space  $(\Omega, (\mathcal{F}_t)_{t \in [0, T]}, \mathbf{P})$  with the filtration  $(\mathcal{F}_t)_{t \in [0, T]}$  satisfying the *usual conditions*. Define the stochastic differential equation (called *forward*)

$$dX_t = S(t, X_t) dt + \sigma(t, X_t) dW_t, \quad X_0, \quad 0 \leq t \leq T,$$

where  $(W_t, \mathcal{F}_t, 0 \leq t \leq T)$  is the standard Wiener process and  $X_0$  is  $\mathcal{F}_0$  measurable initial value. The trend coefficient  $S(t, x)$  and diffusion coefficient  $\sigma^2(t, x)$  satisfy the Lipschitz and linear growth conditions

$$|S(t, x) - S(t, y)|^2 + |\sigma(t, x) - \sigma(t, y)|^2 \leq L|x - y|^2, \quad (2.2)$$

$$S^2(t, x) + \sigma^2(t, x) \leq C(1 + x^2), \quad (2.3)$$

for all  $x, y \in \mathbf{R}$  and for all  $t \in [0, T]$ . Here  $L > 0$  and  $C > 0$  are some constants. By these conditions the stochastic differential equation has a unique strong solution (see Liptser and Shiryaev [29]).

Further, for given two functions  $f(t, x, y, z)$  and  $\Phi(x)$  we have to construct a couple of processes  $(Y_t, Z_t)$  such that the solution of the stochastic differential equation

$$dY_t = -f(t, X_t, Y_t, Z_t) dt + Z_t dW_t, \quad 0 \leq t \leq T,$$

(called *backward*) has the terminal value  $Y_T = \Phi(X_T)$ .

This equation is often written as follows

$$Y_t = \Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s) ds - \int_t^T Z_s dW_s, \quad 0 \leq t \leq T.$$

We suppose that the functions  $f(t, x, y, z)$  and  $\Phi(x)$  satisfy the conditions

$$\begin{aligned} |f(t, x, y_1, z_1) - f(t, x, y_2, z_2)| &\leq L(|y_1 - y_2| + |z_1 - z_2|), \\ |f(t, x, y, z)| + |\Phi(x)| &\leq C(1 + |x|^p), \end{aligned}$$

for all  $x, y, z, y_i, z_i \in \mathbf{R}$ ,  $i = 1, 2$  and for all  $t \in [0, T]$ . Here  $p > 0$ .

This is the so-called *Markovian case*. For the existence and uniqueness of a solution see Pardoux and Peng [34].

The solution  $(Y_t, Z_t)$  can be constructed as follows. Suppose that  $u(t, x)$  satisfies the equation

$$\frac{\partial u}{\partial t} + S(t, x) \frac{\partial u}{\partial x} + \frac{\sigma(t, x)^2}{2} \frac{\partial^2 u}{\partial x^2} = -f\left(t, x, u, \sigma(t, x) \frac{\partial u}{\partial x}\right),$$

with the terminal condition  $u(T, x) = \Phi(x)$ .

Let us put  $Y_t = u(t, X_t)$ , then we obtain by Itô's formula

$$dY_t = \left[ \frac{\partial u}{\partial t} + S(t, X_t) \frac{\partial u}{\partial x} + \frac{\sigma(t, X_t)^2}{2} \frac{\partial^2 u}{\partial x^2} \right] dt + \sigma(t, X_t) \frac{\partial u}{\partial x} dW_t.$$

Hence if we denote  $Z_t = \sigma(t, X_t) u'_x(t, X_t)$  then this equation becomes

$$dY_t = -f(t, X_t, Y_t, Z_t) dt + Z_t dW_t, \quad Y_0 = u(0, X_0)$$

with the terminal value  $Y_T = u(T, X_T) = \Phi(X_T)$ .

### 2.2.2 Black and Scholes model

To illustrate the results from the previous section we consider a particular model, which is called the Black-Scholes model. We consider the following forward equation

$$dX_t = \alpha X_t dt + \vartheta X_t dW_t, \quad X_0 = x_0, \quad 0 \leq t \leq T$$

and two functions  $f(x, y, z) = \beta y + \gamma x z$  and  $\Phi(x)$ . We have to approximate the solution of the backward equation

$$dY_t = -\beta Y_t dt - \gamma X_t Z_t dt + Z_t dW_t, \quad Y_T = \Phi(X_T)$$

in the situation where  $\vartheta \in (a, b)$ ,  $a > 0$  is unknown.

The corresponding partial differential equation is

$$\frac{\partial u}{\partial t} + (\alpha + \vartheta \gamma) x \frac{\partial u}{\partial x} + \frac{\vartheta^2 x^2}{2} \frac{\partial^2 u}{\partial x^2} + \beta u = 0, \quad u(T, x, \vartheta) = \Phi(x).$$

The solution of this equation is the function

$$u(t, x, \vartheta) = \frac{e^{\beta(T-t)}}{\sqrt{2\pi\vartheta^2(T-t)}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2\vartheta^2(T-t)}} \Phi\left(x e^{(\alpha+\vartheta\gamma-\frac{\vartheta^2}{2})(T-t)-z}\right) dz.$$

Hence, using this function we can find the solution of the FBSDE.

## 2.3 Statement of the problem

### 2.3.1 Continuous time observations

Let us remind the situation which we have in the case of continuous time observations of the solution of the stochastic differential equation

$$dX_t = S(t, X_t) dt + \sigma(\vartheta, t, X_t) dW_t, \quad X_0, \quad 0 \leq t \leq T.$$

(see, e.g., [41]). By Itô's formula

$$X_t^2 = X_0^2 + 2 \int_0^t X_s dX_s + \int_0^t \sigma(\vartheta_0, s, X_s)^2 ds,$$

where  $\vartheta_0$  is the true value.

The trajectory fitting estimator (TFE)  $\vartheta^*$  of the parameter  $\vartheta$  can be defined as follows

$$\begin{aligned} \inf_{\vartheta \in \Theta} \int_0^T \left[ X_t^2 - X_0^2 - 2 \int_0^t X_s dX_s - \int_0^t \sigma^2(\vartheta, s, X_s) ds \right]^2 dt \\ = \int_0^T \left[ X_t^2 - X_0^2 - 2 \int_0^t X_s dX_s - \int_0^t \sigma^2(\vartheta^*, s, X_s) ds \right]^2 dt. \end{aligned}$$

In this case also the estimator is equal to the unknown parameter (with probability 1)  $\vartheta^* = \vartheta_0$  under the following mild identifiability condition : for any  $\nu > 0$  with probability 1 we have

$$\inf_{|\vartheta - \vartheta_0| > \nu} \int_0^T \left[ \int_0^t \sigma(\vartheta, s, X_s)^2 ds - \int_0^t \sigma(\vartheta_0, s, X_s)^2 ds \right]^2 dt > 0.$$

If this condition is not fulfilled, then on an event of positive probability, for some  $\vartheta_1 \neq \vartheta_0$  we have

$$\int_0^t \sigma(\vartheta_1, s, X_s)^2 ds = \int_0^t \sigma(\vartheta_0, s, X_s)^2 ds, \quad \forall t \in [0, T],$$

which implies that for all  $t \in [0, T]$

$$\sigma^2(\vartheta_1, t, X_t) = \sigma^2(\vartheta_0, t, X_t).$$

In such a situation no estimation method can provide us a consistent estimator.

Let us illustrate this situation by several examples.

**Example 1.** Suppose that  $\sigma_t(\vartheta, x) = \sqrt{\vartheta} h_t(x)$ ,  $\vartheta \in (\alpha, \beta)$ ,  $\alpha > 0$ , and the observed process is

$$dX_t = S_t(X) dt + \sqrt{\vartheta} h_t(X) dW_t, \quad X_0, \quad 0 \leq t \leq T,$$

where  $S_t(X)$  and  $h_t(X)$  are some functionals of the past, say,

$$S_t(X) = M(t, X_t) + \int_0^t N(s, X_s) ds, \quad h_t(X) = P(t, X_t) + \int_0^t q(s, X_s) ds,$$

where  $M(\cdot)$ ,  $N(\cdot)$ ,  $P(\cdot)$ ,  $q(\cdot)$  are smooth functions. This is an example of so-called *diffusion type* process [29].

To estimate  $\vartheta$  without error we use two approaches. The first one is the TFE

$$\begin{aligned} \vartheta^* &= \arg \inf_{\vartheta \in \Theta} \int_0^T \left[ X_t^2 - X_0^2 - 2 \int_0^t X_s dX_s - \vartheta \int_0^t h_s^2(X) ds \right]^2 dt \\ &= D_T(h)^{-1} \int_0^T \left[ X_t^2 - X_0^2 - 2 \int_0^t X_s dX_s \right] \int_0^t h_s^2(X) ds dt = \vartheta_0, \end{aligned}$$



where

$$D_T(h) = \int_0^T \left( \int_0^t h_s^2(X) ds \right)^2 dt$$

The second possibility is the following. Let  $G(x)$  be a two-times continuously differentiable function. By the Itô's formula for  $G(X_t)$  we can write

$$G(X_t) = G(X_0) + \int_0^t G'(X_s) dX_s + \frac{\vartheta_0}{2} \int_0^t G''(X_s) h_s^2(X) ds.$$

We solve this equation w.r.t. (with respect to)  $\vartheta_0$  and obtain for all  $t \in (0, T]$  with probability 1 the equality

$$\bar{\vartheta}_t = \frac{2G(X_t) - 2G(X_0) - 2 \int_0^t G'(X_s) dX_s}{\int_0^t G''(X_s) h_s^2(X) ds} = \vartheta_0.$$

Therefore we have for all  $t \in (0, T]$  the estimator  $\bar{\vartheta}_t = \vartheta_0$ . Note that we need not to know  $S(\cdot)$  and the only condition we use is that for all  $t \in (0, T]$

$$\int_0^t G''(X_s) h_s^2(X) ds \neq 0.$$

Therefore we obtain an estimator of the unknown parameter without error.

**Example 2.** Suppose that the unknown parameter is  $\vartheta = (\vartheta_1, \dots, \vartheta_d) \in \mathbf{R}_+^d$  and the diffusion coefficient

$$\sigma^2(\vartheta, t, X_t) = \lambda + \sum_{l=1}^d \vartheta_l h_l(t, X_t), \quad \vartheta_l \in (\alpha, \beta),$$

where  $\lambda > 0$  and the functions  $h_l(\cdot) > 0$  are known and  $\vartheta_l > 0$ ,  $l = 1, \dots, d$ . The observed diffusion process is

$$dX_t = S(t, X_t) dt + \sigma(\vartheta, t, X_t) dW_t, \quad X_0, \quad 0 \leq t \leq T.$$

If we denote the vector  $H_t = (H_{1,t}, \dots, H_{d,t})$

$$H_{l,t} = \int_0^t h_l(s, X_s) ds, \quad l = 1, \dots, d,$$

and introduce  $d \times d$  matrix  $\mathbb{H}_t$  and vector  $\tilde{X}_t$  by the relations

$$\mathbb{H}_t = \int_0^t H_s H_s^\top ds, \quad \tilde{X}_t = \left[ X_t^2 - X_0^2 - 2 \int_0^t X_s dX_s - \lambda t \right] H_t,$$

then the TFE is

$$\vartheta_t^* = \mathbb{H}_t^{-1} \int_0^t \left[ X_s^2 - X_0^2 - 2 \int_0^s X_v dX_v - \lambda s \right] H_s ds = \vartheta_0.$$

Here we suppose that the matrix  $\mathbb{H}_t$  for some values of  $t$  is non degenerate and we calculate the estimator for these values of  $t$ . We see that once more we estimate the unknown parameter without error.

Therefore in the case of continuous time observations the approximations

$$\hat{Y}_t = u(t, X_t, \bar{\vartheta}_t) \quad \text{and} \quad \hat{Z}_t = u'_x(t, X_t, \bar{\vartheta}_t) \sigma(\bar{\vartheta}_t, t, X_t)$$

or  $\hat{Y}_t = u(t, X_t, \vartheta_t^*)$  and  $\hat{Z}_t = u'_x(t, X_t, \vartheta_t^*) \sigma(\vartheta_t^*, t, X_t)$  are without errors:  $\hat{Y}_t = Y_t$ ,  $\hat{Z}_t = Z_t$ .

### 2.3.2 High frequency asymptotics

Recall that in the case of continuous time observations there is no statistical problem of estimation of  $\vartheta$  because the measures  $\{\mathbf{P}_\vartheta^{(T)}, \vartheta \in \Theta\}$  corresponding to different values of  $\vartheta$  are singular.

Then in Example 1 with  $h_t(X) = h(t, X_t)$  and  $G(x) = x^2$  we obtain the well-known estimator

$$\bar{\vartheta}_{k,n} = \frac{X_{t_k}^2 - X_0^2 - 2 \sum_{j=1}^k X_{t_{j-1}} (X_{t_j} - X_{t_{j-1}})}{\sum_{j=1}^k h(t_{j-1}, X_{t_{j-1}})^2 \delta}, \quad \delta = \frac{T}{n}.$$

It can be easily shown that if  $n \rightarrow \infty$  then for a fixed  $t$  and corresponding  $k = \lfloor \frac{nt}{T} \rfloor$  we have these convergences in probability

$$\begin{aligned} \sum_{j=1}^k X_{t_{j-1}} (X_{t_j} - X_{t_{j-1}}) &\longrightarrow \int_0^t X_s dX_s, \\ \sum_{j=1}^k h(t_{j-1}, X_{t_{j-1}})^2 \delta &\longrightarrow \int_0^t h(s, X_s)^2 ds \end{aligned}$$

and therefore, in probability,

$$\bar{\vartheta}_{k,n} \longrightarrow \frac{X_t^2 - X_0^2 - 2 \int_0^t X_s dX_s}{\int_0^t h(s, X_s)^2 ds} = \vartheta_0.$$

## 2.4 Parameter estimation

### 2.4.1 Minimum contrast estimator

In this section we are considering the construction of the estimators in general case. The observed forward equation

$$dX_t = S(t, X_t) dt + \sigma(\vartheta, t, X_t) dW_t, \quad X_0, \quad 0 \leq t \leq T \quad (2.4)$$

can be written as

$$X_{t_j} - X_{t_{j-1}} = \int_{t_{j-1}}^{t_j} S(s, X_s) ds + \int_{t_{j-1}}^{t_j} \sigma(\vartheta, s, X_s) dW_s. \quad (2.5)$$

but we consider a (wrong) model which we obtain if we replace the functions  $S(s, X_s)$  and  $\sigma(\vartheta, s, X_s)$  in these integrals by the constant on the interval  $[t_{j-1}, t_j]$  values  $S(t_{j-1}, X_{t_{j-1}})$  and  $\sigma(\vartheta, t_{j-1}, X_{t_{j-1}})$  respectively. Then we obtain

$$X_{t_j} - X_{t_{j-1}} = S(t_{j-1}, X_{t_{j-1}}) \delta + \sigma(\vartheta, t_{j-1}, X_{t_{j-1}}) (W_{t_j} - W_{t_{j-1}}). \quad (2.6)$$

Note that if (2.6) is true then the random variables

$$\frac{X_{t_j} - X_{t_{j-1}} - S(t_{j-1}, X_{t_{j-1}}) \delta}{\sigma(\vartheta, t_{j-1}, X_{t_{j-1}}) \sqrt{\delta}} \quad j = 1, \dots, n$$

are i.i.d. with the standard Gaussian distribution  $\mathcal{N}(0, 1)$ .

Introduce the log pseudo-likelihood for the model (2.6)

$$\begin{aligned} L_{t,k}(\vartheta, X^k) &= -\frac{1}{2} \sum_{j=0}^k \ln [2\pi\sigma^2(\vartheta, t_{j-1}, X_{t_{j-1}}) \delta] \\ &\quad - \sum_{j=1}^k \frac{[X_{t_j} - X_{t_{j-1}} - S(t_{j-1}, X_{t_{j-1}}) \delta]^2}{2\sigma^2(\vartheta, t_{j-1}, X_{t_{j-1}}) \delta} \end{aligned}$$

and define the pseudo-maximum likelihood estimator (PMLE)  $\hat{\vartheta}_{t,n}$  by the equation

$$L_{t,k}(\hat{\vartheta}_{t,n}, X^k) = \sup_{\theta \in \Theta} L_{t,k}(\theta, X^k), \quad .$$

As it was already explained such estimator cannot be used for the construction of the approximations of BSDE due to the complexity of the calculations of the solution of this equation for all  $k$  in nonlinear case. Below we will use this estimator as a preliminary one for the construction of an one-step MLE-process.

*Regularity conditions. (R)*

$R_1$  The functions  $S(\cdot)$  and  $\sigma(\cdot)$  satisfy the conditions of Lipschitz and of linear growth.

$R_2$  The function  $\sigma(\vartheta, t, x)$  is strictly positive and has two continuous derivatives w.r.t.  $\vartheta$ .

$$|\dot{\sigma}(\vartheta, t, x)| + \|\ddot{\sigma}(\vartheta, t, x)\| \leq C(1 + |x|^p) \quad (2.7)$$

$R_3$  With probability one, the information matrix

$$\mathbb{I}_t(\vartheta) = 2 \int_0^t \frac{\dot{\sigma}(\vartheta, s, X_s) \dot{\sigma}(\vartheta, s, X_s)^\top}{\sigma^2(\vartheta, s, X_s)} ds$$

is strictly positive for all  $t \in (0, T]$ .

$R_4$  The function  $u(t, x, \vartheta)$  is continuously differentiable w.r.t.  $\vartheta$  and the derivative satisfies the condition

$$|\dot{u}(t, x, \vartheta)| \leq C(1 + |x|^p).$$

It is convenient to replace the likelihood ratio function by the contrast function

$$\begin{aligned} U_{t,k}(\vartheta, X^k) &= \sum_{j=1}^k \delta \ln a(\vartheta, t_{j-1}, X_{t_{j-1}}) \\ &\quad + \sum_{j=1}^k \frac{(X_{t_j} - X_{t_{j-1}} - S(t_{j-1}, X_{t_{j-1}}) \delta)^2}{a(\vartheta, t_{j-1}, X_{t_{j-1}})}, \end{aligned}$$

where  $a(\vartheta, t, x) = \sigma(\vartheta, t, x)^2$ . The estimator  $\hat{\vartheta}_{t,n}$  satisfies the equation

$$U_{t,k}(\hat{\vartheta}_{t,n}, X^k) = \inf_{\vartheta \in \Theta} U_{t,k}(\vartheta, X^k). \quad (2.8)$$

The contrast function converges to the following limit

$$\begin{aligned} U_{t,k}(\vartheta, X^k) &\longrightarrow U_t(\vartheta, X^t) = \int_0^t \left[ \frac{a(\vartheta_0, s, X_s)}{a(\vartheta, s, X_s)} - \ln \frac{a(\vartheta_0, s, X_s)}{a(\vartheta, s, X_s)} \right] ds \\ &\quad + \int_0^t \ln a(\vartheta_0, s, X_s) ds. \end{aligned}$$

*Identifiability condition.*

$I_1$  The random function  $U_t(\vartheta, X^t)$ ,  $\vartheta \in \Theta$ ,  $0 < t \leq T$  with probability 1 has a unique minimum at the point  $\vartheta = \vartheta_0$

$$\inf_{\vartheta \in \Theta} U_t(\vartheta, X^t) = U_t(\vartheta_0, X^t), \quad 0 < t \leq T.$$

In reality this condition requires only the uniqueness of the solution since the function  $x \mapsto x - \ln x$  attains its minimum at  $x = 1$  implies that  $U_t(\vartheta, X^t)$  is always larger than  $U_t(\vartheta_0, X^t)$ . Hence  $\vartheta_0$  is always a minimum point of the mapping  $\vartheta \mapsto U_t(\vartheta, X^t)$ .

Introduce the vector-process

$$\xi_t(\vartheta_0) = \mathbb{I}_t(\vartheta_0)^{-1} \sqrt{2} \int_0^t \frac{\dot{\sigma}(\vartheta_0, s, X_s)}{\sigma(\vartheta_0, s, X_s)} dw(s), \quad 0 < t \leq T.$$

Note that the Wiener process  $w(s)$ ,  $0 \leq s \leq T$  here is independent on the diffusion process  $X_s$ ,  $0 \leq s \leq T$ .

For given  $t \in (0, T]$  the value  $t_k$  in the estimator  $\hat{\vartheta}_{t_k, n}$  satisfies the condition  $t_k \leq t < t_{k+1}$ .

## 2.4.2 Local asymptotic mixed normality

**Theorem 2.1.** *Suppose that the Regularity and Identifiability conditions are fulfilled. Then for all  $t \in (0, T]$  the estimator  $\hat{\vartheta}_{t_k, n}$  is consistent and asymptotically conditionally normal (stable convergence)*

$$\sqrt{\frac{n}{T}} \left( \hat{\vartheta}_{t_k, n} - \vartheta_0 \right) \Longrightarrow \xi_t(\vartheta_0). \quad (2.9)$$

Moreover this estimator is asymptotically efficient.

The proofs of this theorem can be found in [4] (lower bound,  $d = 1$ ) and in [9] (properties of estimator,  $d \geq 1$ ).

Let us outline the proof. Suppose that the consistency of the estimator  $\hat{\vartheta}_{t_k, n}$  defined by the equation (2.8) is already proved.

Introduce independent random variables

$$w_j = (2\delta)^{-1/2} \left[ (W_{t_j} - W_{t_{j-1}})^2 - \delta \right], \quad \mathbf{E}w_j = 0, \quad \mathbf{E}w_j^2 = \delta, \quad \mathbf{E}w_j w_i = 0$$

for  $j \neq i$  and note that the empirical Fisher information matrix

$$\mathbb{I}_{t,n}(\vartheta_0) = 2 \sum_{j=1}^k \frac{\dot{\sigma}(\vartheta_0, t_{j-1}, X_{t_{j-1}}) \dot{\sigma}(\vartheta_0, t_{j-1}, X_{t_{j-1}})^\top}{\sigma^2(\vartheta_0, t_{j-1}, X_{t_{j-1}})} \delta \longrightarrow \mathbb{I}_t(\vartheta_0) \quad (2.10)$$

as  $n \rightarrow \infty$ . Then by the Taylor's expansion of the solution  $\hat{\vartheta}_{t,n}$  of the system of  $d$ -equations

$$\frac{\partial U_{t,k}(\vartheta, X^k)}{\partial \vartheta} = 0$$

we can write the representation of the MCE

$$\begin{aligned} \delta^{-1/2} \left( \hat{\vartheta}_{t,n} - \vartheta_0 \right) &= \\ \mathbb{I}_{t,n}(\vartheta_0)^{-1} \sqrt{2} \sum_{j=1}^k \frac{\dot{\sigma}(\vartheta_0, t_{j-1}, X_{t_{j-1}})}{\sigma(\vartheta_0, t_{j-1}, X_{t_{j-1}})} w_j (1 + o(1)). \end{aligned}$$

The symbols  $o, O$  are always understood in the sense of convergence in probability.

Now the convergence (2.9) follows from (2.10) and (stable convergence)

$$\sum_{j=1}^k \frac{\dot{\sigma}(\vartheta_0, t_{j-1}, X_{t_{j-1}})}{\sigma(\vartheta_0, t_{j-1}, X_{t_{j-1}})} w_j \Longrightarrow \int_0^t \frac{\dot{\sigma}(\vartheta_0, s, X_s)}{\sigma(\vartheta_0, s, X_s)} dw(s).$$

Note that the approximation  $\hat{Y}_{t,n} = u(t, X_{t_k}, \hat{\vartheta}_{t,n})$  is computationally difficult to realize because solving the equation (2.8) for all  $t_k, k = 1, \dots, n$  especially in non linear case is almost impossible. This is the reason for proposing the one-step MLE-process as follows.

Let us fix some (small)  $\tau > 0$  and denote by  $\hat{\vartheta}_{\tau,n}$  the MCE constructed by the observations  $X^{\tau,n} = (X_0, X_{t_{1,n}}, \dots, X_{t_{N,n}})$ , where  $t_{N,n} \leq \tau < t_{N+1,n}$ .

By the Theorem 2.1, this estimator is consistent and asymptotically conditionally normal

$$\begin{aligned} \sqrt{\frac{n}{T}} \left( \hat{\vartheta}_{\tau,n} - \vartheta_0 \right) &= \mathbb{I}_{\tau,n}(\vartheta_0)^{-1} \sqrt{2} \sum_{j=1}^N \frac{\dot{\sigma}(\vartheta_0, t_{j-1}, X_{t_{j-1}})}{\sigma(\vartheta_0, t_{j-1}, X_{t_{j-1}})} w_j + o(1) \\ &\Longrightarrow \xi_\tau(\vartheta_0) = \mathbb{I}_\tau(\vartheta_0)^{-1} \sqrt{2} \int_0^\tau \frac{\dot{\sigma}(\vartheta_0, s, X_s)}{\sigma(\vartheta_0, s, X_s)} dw(s). \end{aligned}$$

Here the random Fisher information matrix is

$$\mathbb{I}_\tau(\vartheta_0) = 2 \int_0^\tau \frac{\dot{\sigma}(\vartheta_0, s, X_s) \dot{\sigma}(\vartheta_0, s, X_s)^\top}{\sigma^2(\vartheta_0, s, X_s)} ds.$$

## 2.5 Approximation of the solution

In this section we are considering a statistical problem of the estimation of the solution of a FBSDE. We suppose that the diffusion coefficient of the forward equation depends on an unknown parameter, hence the solution of the FBSDE also depends on that unknown parameter. Based on the observations from the solution of the forward equation we construct two process-estimators to efficiently approximate the solution of the FBSDE.

### 2.5.1 Estimator process

We consider the problem of the approximation of the solution  $(Y_t, Z_t)$  of BSDE in the situations, where the forward equation contains an unknown finite-dimensional parameter  $\vartheta$ :

$$dX_t = S(t, X_t) dt + \sigma(\vartheta, t, X_t) dW_t, \quad X_0, \quad 0 \leq t \leq T. \quad (2.11)$$

Then the solution  $u$  of the corresponding partial differential equation

$$\frac{\partial u}{\partial t} + S(t, x) \frac{\partial u}{\partial x} + \frac{\sigma^2(\vartheta, t, x)}{2} \frac{\partial^2 u}{\partial x^2} = -f\left(t, x, u, \sigma(\vartheta, t, x) \frac{\partial u}{\partial x}\right),$$

depends on  $\vartheta$ , i.e.,  $u = u(t, x, \vartheta)$ . The backward equation

$$dY_t = -f(t, X_t, Y_t, Z_t) dt + Z_t dW_t, \quad 0 \leq t \leq T,$$

we obtain if we put  $Y_t = u(t, X_t, \vartheta)$  and  $Z_t = u'_x(t, X_t, \vartheta) \sigma(\vartheta, t, X_t)$ . But as  $\vartheta$  is unknown we propose the natural approximations

$$\hat{Y}_t = u(t, X_t, \vartheta_t^*), \quad \hat{Z}_t = u'_x(t, X_t, \vartheta_t^*) \sigma(\vartheta_t^*, t, X_t).$$

Here  $\vartheta_t^*, 0 \leq t \leq T$  is some *good estimator-process* of  $\vartheta$  with *small error*. In this problem the good estimator means the following

- $\vartheta_t^* = \vartheta_t^*(X^t)$ , i.e., *it depends on* observations  $X^t = (X_s, 0 \leq s \leq t)$  till time  $t$ .
- *Easy to calculate for each*  $t \in (0, T]$ .
- *Provides an asymptotically efficient estimator of*  $Y_t$ , i.e., *we have in some sense*

$$\mathbf{E}_\vartheta \left( \hat{Y}_t - Y_t \right)^2 \rightarrow \min_{\bar{Y}_t} \mathbf{E}_\vartheta \left( \bar{Y}_t - Y_t \right)^2.$$

As we have already fixed the approximation  $Y_t$  as  $\hat{Y}_t = u(t, X_t, \vartheta_t^*)$  the main problem is *how to find a good estimator-process*  $\vartheta_t^*, 0 \leq t \leq T$ ?

Observe that the problem of estimation of  $\vartheta$  is singular, i.e., the parameter  $\vartheta$  can be estimated by continuous time observations without error. The problem become more interesting if we consider the discrete time observations. Suppose that the solution of the equation (2.4) is observed in discrete times  $t_i = i\frac{T}{n}$  and we have to study the approximations

$$\hat{Y}_{t_k} = u(t_k, X_{t_k}, \hat{\vartheta}_{t_k}), \quad \hat{Z}_{t_k} = \sigma \left( \hat{\vartheta}_{t_k}, t_k, X_{t_k} \right) u'_x(t_k, X_{t_k}, \hat{\vartheta}_{t_k}), k = 1, \dots, n,$$

of the solution  $Y_t, Z_t$  of BSDE

$$dY_t = -f(t, X_t, Y_t, Z_t) dt + Z_t dW_t, \quad 0 \leq t \leq T.$$

Here  $k$  satisfies the conditions  $t_k \leq t < t_{k+1}$  and the estimator  $\hat{\vartheta}_{t_k}$  can be constructed by the observations  $X^k = (X_0, X_{t_1}, \dots, X_{t_k})$  up to time  $t_k$ . The properties of estimators we study in the so-called *higher frequency* asymptotics:  $n \rightarrow \infty$ . Observe that the problem of estimation of the parameter  $\vartheta$  in the case of discrete-time observations of the processes like (2.4) was extensively studied last years (see, e.g., [38] and the references therein).

## 2.5.2 One-step maximum likelihood estimator (MLE) process

Introduce the pseudo score-function  $(A_{j-1}(\vartheta) = \sigma(\vartheta, t_{j-1}, X_{t_{j-1}})^2)$

$$\begin{aligned} \Delta_{k,n}(\vartheta, X^k) &= \sum_{j=1}^k \dot{\ell}(\vartheta, X_{t_{j-1}}, X_{t_j}) \\ &= \sum_{j=1}^k \frac{\left[ (X_{t_j} - X_{t_{j-1}} - S_{j-1} \delta)^2 - A_{j-1}(\vartheta) \delta \right] \dot{A}_{j-1}(\vartheta)}{2A_{j-1}^2(\vartheta) \sqrt{\delta}}. \end{aligned}$$

For any  $t \in [\tau, T]$  define  $k$  by the condition  $t_k \leq t < t_{k+1}$  and the one-step PMLE-process by the relation

$$\vartheta_{k,n}^* = \hat{\vartheta}_{\tau,n} + \sqrt{\delta} \mathbb{I}_{k,n}(\hat{\vartheta}_{\tau,n})^{-1} \Delta_{k,n}(\hat{\vartheta}_{\tau,n}, X^k), \quad k = N + 1, \dots, n. \quad (2.12)$$

Our goal is to show that the corresponding approximation

$$Y_{t_k,n}^* = u(t_k, X_{t_k}, \vartheta_{k,n}^*), k = N + 1, \dots, n$$



is asymptotically efficient. To do this we need to present the lower bound on the risks of all estimators and then to show that for the proposed approximation this lower bound is achieved.

Of course we could construct the approximation of  $Y_t$  and  $Z_t$  as follows

$$\hat{Y}_{t_k,n} = u(t_k, X_{t_k}, \hat{\vartheta}_{k,n}) \text{ and } \hat{Z}_{t_k,n} = u'_x(t_k, X_{t_k}, \hat{\vartheta}_{k,n})\sigma\left(\hat{\vartheta}_{k,n}, t_k, X_{t_k}\right),$$

that is, using only the first estimator. Note that this type of approximation is not asymptotically efficient, since for constructing the estimator for the parameter we are not using whole observations, but only part of it. That is why we are constructing another estimator of  $\vartheta$  which can provide smaller error of estimation.

### 2.5.3 Local asymptotic mixed normality

First we recall some known results on asymptotically efficient estimation of the parameter  $\vartheta$ . The family of measures  $\{\mathbf{P}_{\vartheta}^{(k,n)}, \vartheta \in \Theta\}$  induced in  $R^k$  by the observations  $X^k$  are *locally asymptotically mixed normal* (LAMN), i.e., the likelihood ratio

$$Z_{k,n}(v) = \frac{d\mathbf{P}_{\vartheta_0 + \frac{v}{\sqrt{n}}}^{(k,n)}}{d\mathbf{P}_{\vartheta_0}^{(k,n)}}, \quad v \in V_n = \left\{v : \vartheta_0 + \frac{v}{\sqrt{n}} \in \Theta\right\},$$

admits the representation

$$Z_{k,n}(v) = \exp \left\{ \langle v, \Delta_{k,n}(\vartheta_0, X^k) \rangle - \frac{1}{2} v \mathbb{I}_{k,n}(\vartheta_0) v^\top + r_n \right\},$$

where  $r_n \rightarrow 0 = r_n(v, \vartheta_0) \rightarrow 0$  in probability for fixed  $\vartheta_0 \in \Theta$  and fixed  $v \in \mathbf{R}$ . The proof can be found in [4] ( $d = 1$ ) and in [11] ( $d \geq 1$ ).

In statistical problems with such property of the families of measures we have, so-called, Jeganathan-type lower bound on the risks of all estimators  $\bar{\vartheta}_{k,n}$ :

$$\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \sup_{|\vartheta - \vartheta_0| \leq \varepsilon} \mathbf{E}_{\vartheta} \ell(\delta^{-1/2}(\bar{\vartheta}_{k,n} - \vartheta)) \geq \mathbf{E}_{\vartheta_0} \ell(\xi_t(\vartheta_0)).$$

Here  $\ell(v)$ ,  $v \in R^d$  is some symmetric, non decreasing loss function (see the conditions in [16]).

Therefore we can call estimator  $\vartheta_{k,n}^*$  is asymptotically efficient if for some function  $\ell(\cdot)$  and all  $\vartheta_0 \in \Theta$  we have the equality

$$\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \sup_{|\vartheta - \vartheta_0| \leq \varepsilon} \mathbf{E}_{\vartheta} \ell(\delta^{-1/2}(\vartheta_{k,n}^* - \vartheta)) = \mathbf{E}_{\vartheta_0} \ell(\xi_t(\vartheta_0)).$$

We say that the estimator-process  $\vartheta_{k,n}^*$ ,  $k = N + 1, \dots, n$  is asymptotically efficient for the values  $t \in [\tau_*, T]$ , if we have this equality for all  $t \in [\tau_*, T]$ . Here  $0 < \tau < \tau_* < T$

**Theorem 2.2.** *The one-step MLE-process  $\vartheta_{k,n}^*$ ,  $k = N + 1, \dots, n$  is consistent, asymptotically conditionally normal (stable convergence)*

$$\delta^{-1/2} (\vartheta_{k,n}^* - \vartheta_0) \implies \xi_t(\vartheta_0) \quad (2.13)$$

and is asymptotically efficient for  $t \in [\tau_*, T]$  where  $\tau < \tau_* < T$  and a bounded loss functions.

*Proof.* The proof follows the main steps of the similar proof given in [23].

We have for any  $\nu > 0$  the estimates

$$\begin{aligned} \mathbf{P}_{\vartheta_0}^{(k,n)} \left\{ |\vartheta_{k,n}^* - \vartheta_0| > \nu \right\} &\leq \mathbf{P}_{\vartheta_0}^{(k,n)} \left\{ \left| \hat{\vartheta}_{\tau,n} - \vartheta_0 \right| > \frac{\nu}{2} \right\} \\ &+ \mathbf{P}_{\vartheta_0}^{(k,n)} \left\{ \left| \sqrt{\delta} \mathbb{I}_{k,n}(\hat{\vartheta}_{\tau,n})^{-1} \Delta_{k,n}(\hat{\vartheta}_{\tau,n}, X^k) \right| > \frac{\nu}{2} \right\}. \end{aligned}$$

We can write

$$\left| \mathbb{I}_{k,n}(\hat{\vartheta}_{\tau,n}) - \mathbb{I}_{k,n}(\vartheta_0) \right| \leq C \left| \hat{\vartheta}_{\tau,n} - \vartheta_0 \right| \longrightarrow 0$$

and

$$\sqrt{\delta} \left| \Delta_{k,n}(\hat{\vartheta}_{\tau,n}, X^k) - \Delta_{k,n}(\vartheta_0, X^k) \right| \longrightarrow 0.$$

Further, it can be shown that

$$\mathbf{P}_{\vartheta_0}^{(k,n)} \left\{ \left| \sqrt{\delta} \mathbb{I}_{k,n}(\vartheta_0)^{-1} \Delta_{k,n}(\vartheta_0, X^k) \right| > \frac{\nu}{2} \right\}.$$

Moreover, more detailed analysis allows to verify the uniform consistency as well:

$$\mathbf{P}_{\vartheta_0}^{(k,n)} \left\{ \max_{N+1 \leq k \leq n} |\vartheta_{k,n}^* - \vartheta_0| > \nu \right\} \longrightarrow 0$$

See the similar problem in [23], Theorem 1. The asymptotic conditional normality as well follows from the similar steps. We have

$$\begin{aligned} \delta^{-1/2} (\vartheta_{k,n}^* - \vartheta_0) &= \delta^{-1/2} (\hat{\vartheta}_{\tau,n} - \vartheta_0) + \mathbb{I}_{k,n}(\hat{\vartheta}_{\tau,n})^{-1} \Delta_{k,n}(\hat{\vartheta}_{\tau,n}, X^k) \\ &= \delta^{-1/2} (\hat{\vartheta}_{\tau,n} - \vartheta_0) + \mathbb{I}_{k,n}(\hat{\vartheta}_{\tau,n})^{-1} \Delta_{k,n}(\vartheta_0, X^k) \\ &\quad + \mathbb{I}_{k,n}(\hat{\vartheta}_{\tau,n})^{-1} \left[ \Delta_{k,n}(\hat{\vartheta}_{\tau,n}, X^k) - \Delta_{k,n}(\vartheta_0, X^k) \right]. \end{aligned}$$

The central statistics

$$\mathbb{I}_{k,n}(\hat{\vartheta}_{\tau,n})^{-1} \Delta_{k,n}(\vartheta_0, X^k) \implies \xi_t(\vartheta_0).$$

We have to show that

$$b_n = \delta^{-1/2} \left( \hat{\vartheta}_{\tau,n} - \vartheta_0 \right) + \mathbb{I}_{k,n}(\hat{\vartheta}_{\tau,n})^{-1} \left[ \Delta_{k,n}(\hat{\vartheta}_{\tau,n}, X^k) - \Delta_{k,n}(\vartheta_0, X^k) \right] \longrightarrow 0.$$

The representation

$$\begin{aligned} & \Delta_{k,n}(\hat{\vartheta}_{\tau,n}, X^k) - \Delta_{k,n}(\vartheta_0, X^k) \\ &= \sum_{j=1}^k \int_0^1 \dot{\ell} \left( \vartheta_0 + v \left( \hat{\vartheta}_{\tau,n} - \vartheta_0 \right), X_{t_{j-1}}, X_{t_j} \right) \left( \hat{\vartheta}_{\tau,n} - \vartheta_0 \right) dv \end{aligned}$$

allows us to write

$$\begin{aligned} & \mathbb{I}_{k,n}(\hat{\vartheta}_{\tau,n}) b_n \\ &= \left[ \mathbb{I}_{k,n}(\hat{\vartheta}_{\tau,n}) + \sum_{j=1}^k \int_0^1 \ddot{\ell}(\vartheta(v), X_{t_{j-1}}, X_{t_j}) dv \sqrt{\delta} \right] \frac{\left( \hat{\vartheta}_{\tau,n} - \vartheta_0 \right)}{\sqrt{\delta}}, \end{aligned}$$

where  $\vartheta(v) = \vartheta_0 + v \left( \hat{\vartheta}_{\tau,n} - \vartheta_0 \right)$ . Further

$$\begin{aligned} \sum_{j=1}^k \ddot{\ell}(\vartheta(v), X_{t_{j-1}}, X_{t_j}) &= \sum_{j=1}^k \ddot{\ell}(\vartheta_0, X_{t_{j-1}}, X_{t_j}) + O\left(\hat{\vartheta}_{\tau,n} - \vartheta_0\right) \\ &= - \sum_{j=1}^k \frac{\dot{A}_{j-1}(\vartheta_0) \dot{A}_{j-1}(\vartheta_0)^\top}{2A_{j-1}^2(\vartheta_0)} \sqrt{\delta} + o(1) \end{aligned}$$

because in two other terms after the differentiation

$$\dot{\ell}(\vartheta, X_{t_{j-1}}, X_{t_j}) = \frac{\left[ (X_{t_j} - X_{t_{j-1}} - S_{j-1} \delta)^2 - A_{j-1}(\vartheta) \delta \right] \dot{A}_{j-1}(\vartheta)}{2A_{j-1}^2(\vartheta) \sqrt{\delta}}$$

contains the quantity

$$\begin{aligned} & \left[ X_{t_j} - X_{t_{j-1}} - S_{j-1} \delta \right]^2 - A_{j-1}(\vartheta_0) \delta = -\sigma^2(t_{j-1}, X_{t_{j-1}}, \vartheta_0) \delta \\ & + \left( \int_{t_{j-1}}^{t_j} [S(s, X_s) - S(t_{j-1}, X_{t_{j-1}})] ds + \int_{t_{j-1}}^{t_j} \sigma(s, X_s, \vartheta_0) dW_s \right)^2 \\ &= 2 \int_{t_{j-1}}^{t_j} \tilde{X}_s [S(s, X_s) - S(t_{j-1}, X_{t_{j-1}})] ds + 2 \int_{t_{j-1}}^{t_j} \tilde{X}_s \sigma(s, X_s, \vartheta_0) dW_s \\ & + \int_{t_{j-1}}^{t_j} [\sigma^2(s, X_s, \vartheta_0) - \sigma^2(t_{j-1}, X_{t_{j-1}}, \vartheta_0)] ds = O(\delta^2) + O(\delta). \end{aligned}$$

Here  $\tilde{X}_s = X_s - X_{t_{j-1}} - S_{j-1}s$ . Hence

$$\begin{aligned}\tilde{X}_s &= \int_{t_{j-1}}^s [S(r, X_r) - S(t_{j-1}, X_{t_{j-1}})] dr + \int_{t_{j-1}}^s \sigma(r, X_r, \vartheta_0) dW_r \\ &= O(\delta^{3/2}) + \sigma(t_{j-1}, X_{t_{j-1}}, \vartheta_0) [W_s - W_{t_{j-1}}] + O(\delta).\end{aligned}$$

Note that for the stochastic integral as  $n \rightarrow \infty$  we have

$$\begin{aligned}\int_{t_{j-1}}^{t_j} \tilde{X}_s \sigma(s, X_s, \vartheta_0) dW_s &= \sigma^2(t_{j-1}, X_{t_{j-1}}, \vartheta_0) \int_{t_{j-1}}^{t_j} [W_s - W_{t_{j-1}}] dW_s (1 + o(1)) \\ &= \sigma^2(t_{j-1}, X_{t_{j-1}}, \vartheta_0) \left[ \frac{(W_{t_j} - W_{t_{j-1}})^2 - \delta}{2} \right].\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{I}_{k,n}(\hat{\vartheta}_{\tau,n}) + \sum_{j=1}^k \int_0^1 \ddot{\ell}(\vartheta(v), X_{t_{j-1}}, X_{t_j}) dv \sqrt{\delta} &= \mathbb{I}_{k,n}(\hat{\vartheta}_{\tau,n}) - \sum_{j=1}^k \frac{\dot{A}_{j-1}(\vartheta_0) \dot{A}_{j-1}(\vartheta_0)^\top}{2A_{j-1}^2(\vartheta_0)} \delta + o(1) \\ &= \mathbb{I}_{k,n}(\hat{\vartheta}_{\tau,n}) - \mathbb{I}_{k,n}(\vartheta_0) + o(1) \longrightarrow 0.\end{aligned}$$

The obtained relations prove the convergence (2.13). More detailed analysis shows that this convergence is locally uniform. Hence the one-step MLE-process is asymptotically efficient for the bounded loss functions.  $\blacksquare$

In the next section we are considering asymptotically efficient approximation of the solution of a FBSDE. For that reason we establish a lower bound to compare all estimators.

### 2.5.4 Lower bound on approximation of the solution of a FBSDE

We are considering a system of SDEs,

$$dX_t = S(t, X_t)dt + \sigma(\vartheta, t, X_t)dW_t, \quad X_0 = x_0, \quad t \in [0, T], \quad (2.14)$$

satisfying the conditions of existence and uniqueness. Here the unknown parameter  $\vartheta$  is one-dimensional and belongs to an open interval  $\Theta = (\alpha, \beta) \subset \mathbf{R}$ . We have the observations at discrete times  $0 = t_0 < t_1 < t_2 < \dots < t_n = T$ ,  $t_j = j\frac{T}{n}$  of the solution of this SDE

$$\mathbf{X}^n = (X_{t_0}, X_{t_1}, \dots, X_{t_n}).$$

Denote by  $k$  the number of observation times before the time moment  $t$

$$0 = t_0 < t_1 < t_2 < \dots < t_k \leq t < t_{k+1}, \quad k = \left[ \frac{tn}{T} \right], \quad t_k = k\frac{T}{n}, \quad \delta = \frac{T}{n}.$$

For our statistical model we have only the observations up until the time  $t$

$$\mathbf{X}^k = (X_{t_0}, X_{t_1}, \dots, X_{t_k}),$$

which induce the probability measures  $\mathbf{P}_{\vartheta}^{n,k}$  on  $\mathbf{R}^k$ . The probability measures  $\{\mathbf{P}_{\vartheta}^{n,k}, \vartheta \in \Theta\}$  are equivalent.

**Proposition 2.2** (Gobet, [11]). *If the following conditions are fulfilled*

1. *the diffusion coefficient  $\sigma(\vartheta, t, x)$  is of class  $C^{1+\alpha}$ , ( $\alpha > 0$ ) with respect to the parameter  $\vartheta$ , the functions  $S, \sigma, \dot{\sigma}, S', \sigma'$  are of class  $C^{1,2}$  with respect to  $(t, x)$  and  $\sigma', S'$  are uniformly bounded (dot is the derivative w.r.t.  $\vartheta$ , prime w.r.t. the variable  $x$ ),*
2.  *$\sigma(\vartheta, t, x) > 0$ , for all  $\vartheta \in \Theta$ ,  $x \in \mathbf{R}$ ,  $t \in [0, T]$ ,*

*then the family of probability measures  $\{\mathbf{P}_{\vartheta}^{n,k}, \vartheta \in \Theta\}$  satisfies the LAMN condition at each point  $\vartheta = \vartheta_0$  of the parametric set  $\Theta$ , that is, the Radon-Nikodym derivative of  $\mathbf{P}_{\vartheta_0 + \sqrt{\delta}v}^{n,k}$  w.r.t.  $\mathbf{P}_{\vartheta_0}^{n,k}$  has such a representation*

$$\ln \frac{d\mathbf{P}_{\vartheta_0 + \sqrt{\delta}v}^{n,k}}{d\mathbf{P}_{\vartheta_0}^{n,k}}(\mathbf{X}^k) = v\Delta_{n,k}(\vartheta_0) - \frac{1}{2}v^2\mathbf{I}_{n,k}(\vartheta_0) + r_{n,k}(v, \vartheta_0),$$

*where  $r_{n,k}(v, \vartheta_0) \rightarrow 0$ , in  $\mathbf{P}_{\vartheta_0}^{n,k}$  probability, for every  $v \in \mathbf{R}$  and*

$$\begin{aligned} \Delta_{n,k}(\vartheta_0) &= \sqrt{2} \sum_{j=1}^k \frac{\dot{\sigma}(\vartheta_0, t_{j-1}, X_{t_{j-1}})^2}{\sigma(\vartheta_0, t_{j-1}, X_{t_{j-1}})^2} (w_{t_j} - w_{t_{j-1}}) \implies \Delta_t(\vartheta_0) = \\ &= \sqrt{2} \int_0^t \frac{\dot{\sigma}(\vartheta_0, s, X_s)}{\sigma(\vartheta_0, s, X_s)} dw(s), \end{aligned}$$

*stably in  $\mathbf{P}_{\vartheta_0}^{n,k}$  law,  $w(t)$  is a Wiener process independent on  $\{X_s, 0 \leq s \leq t\}$  and*

$$\mathbf{I}_{n,k}(\vartheta_0) = 2 \sum_{j=1}^k \frac{\dot{\sigma}(\vartheta_0, t_{j-1}, X_{t_{j-1}})^2}{\sigma(\vartheta_0, t_{j-1}, X_{t_{j-1}})^2} \delta \longrightarrow \mathbf{I}_t(\vartheta_0) = 2 \int_0^t \frac{\dot{\sigma}(\vartheta_0, s, X_s)^2}{\sigma(\vartheta_0, s, X_s)^2} ds,$$

*in  $\mathbf{P}_{\vartheta_0}^{n,k}$  probability.*

**Lemma 2.1.** *Under the conditions of Proposition 2.2 the stochastic process*

$$Y_t(\vartheta_0, v) = e^{v\Delta_t(\vartheta_0) - \frac{1}{2}v^2\mathbf{I}_t(\vartheta_0)}$$

*has mathematical expectation equal to 1.*

*Proof.* Denote  $\xi = \frac{\Delta_t(\vartheta_0)}{\sqrt{\mathbf{I}_t(\vartheta_0)}}$ ,  $\eta = \sqrt{\mathbf{I}_t(\vartheta_0)}$ , then  $\xi$  is a standard normal distribution independent on  $\eta$  and

$$Y_t(\vartheta_0, v) = e^{v\xi\eta - \frac{1}{2}v^2\eta^2}.$$

$$\begin{aligned} \mathbf{E}e^{v\xi\eta - \frac{1}{2}v^2\eta^2} &= \int_{\mathbf{R}} \int_{\mathbf{R}} e^{vxy - \frac{1}{2}v^2y^2} d\mathbf{P}(\xi < x, \eta < y) = \\ &= \int_{\mathbf{R}} e^{-\frac{1}{2}v^2y^2} \int_{\mathbf{R}} e^{vxy} e^{-\frac{1}{2}x^2} dx d\mathbf{P}(\eta < y) = \int_{\mathbf{R}} e^{-\frac{1}{2}v^2y^2} e^{\frac{1}{2}v^2y^2} d\mathbf{P}(\eta < y) = 1, \end{aligned}$$

since  $\mathbf{P}(\xi < x, \eta < y) = \mathbf{P}(\xi < x)\mathbf{P}(\eta < y)$ . Hence

$$\mathbf{E}e^{v\Delta_t(\vartheta_0) - \frac{1}{2}v^2\mathbf{I}_t(\vartheta_0)} = 1. \quad \blacksquare$$

**Remark 2.2.** *The LAMN condition does not imply the absolute continuity of measures  $\mathbf{P}_{\vartheta_0 + \sqrt{\delta}v}^{n,k}$  and  $\mathbf{P}_{\vartheta_0}^{n,k}$ , but one can estimate the singular part of these measures by the lemma above. Indeed,*

$$\lim_{n \rightarrow +\infty} \mathbf{E}_{\vartheta_0} \frac{d\mathbf{P}_{\vartheta_0 + \sqrt{\delta}v}^{n,k}}{d\mathbf{P}_{\vartheta_0}^{n,k}}(\mathbf{X}^k) = 1.$$

Since

$$\frac{d\mathbf{P}_{\vartheta_0 + \sqrt{\delta}v}^{n,k}}{d\mathbf{P}_{\vartheta_0}^{n,k}}(\mathbf{X}^k) \implies e^{v\Delta_t(\vartheta_0) - \frac{1}{2}|v|^2\mathbf{I}_t(\vartheta_0)}$$

stably in law, hence

$$\lim_{n \rightarrow +\infty} \mathbf{E}_{\vartheta_0} \frac{d\mathbf{P}_{\vartheta_0 + \sqrt{\delta}v}^{n,k}}{d\mathbf{P}_{\vartheta_0}^{n,k}}(\mathbf{X}^k) \geq \mathbf{E}e^{v\Delta_t(\vartheta_0) - \frac{1}{2}|v|^2\mathbf{I}_t(\vartheta_0)} = 1,$$

on the other hand

$$\mathbf{E}_{\vartheta_0} \frac{d\mathbf{P}_{\vartheta_0 + \sqrt{\delta}v}^{n,k}}{d\mathbf{P}_{\vartheta_0}^{n,k}}(\mathbf{X}^k) \leq 1,$$

finally we obtain

$$\lim_{n \rightarrow +\infty} \mathbf{E}_{\vartheta_0} \frac{d\mathbf{P}_{\vartheta_0 + \sqrt{\delta}v}^{n,k}}{d\mathbf{P}_{\vartheta_0}^{n,k}}(\mathbf{X}^k) = 1. \quad \diamond$$

**Remark 2.3.** *In fact, the expectation is equal to 1 without limit since the measures induced by discrete observations are equivalent.*  $\diamond$

Next proposition is used in proving the lower bound in the statistical model where we have local asymptotic mixed normality property.

**Proposition 2.3** (Jeganathan, [15]). *If the family of measures  $\{\mathbf{P}_\vartheta^{n,k}, \vartheta \in \Theta\}$  defined above satisfies the LAMN condition at the point  $\vartheta = \vartheta_0 \in \Theta$ , then there exist*

- an increasing sequence  $\{m_n\}$  tending to infinity as  $n \rightarrow +\infty$ ,
- functions  $c_n : \Theta \times R \rightarrow R$  so that

$$\sup_{|v| \leq \delta} |c_n(\vartheta_0, v) - 1| \rightarrow 0,$$

as  $n \rightarrow +\infty$ , for every  $\delta > 0$ , such that, denoting

$$\hat{\mathbf{Z}}_{n,k}(v) = c_{n,k}(\vartheta_0, v) e^{[v\Delta_{n,k}^*(\vartheta_0) - \frac{1}{2}v^2\mathbf{I}_{n,k}(\vartheta_0)]},$$

where

$$\Delta_{n,k}^*(\vartheta_0) = \Delta_{n,k}(\vartheta_0) \mathbf{I}(|\Delta_{n,k}(\vartheta_0)| \leq m_n),$$

and

$$\mathbf{Q}_{\vartheta_0, v}^{n,k}(A) = \int_A \hat{\mathbf{Z}}_{n,k}(v) \mathbf{P}_{\vartheta_0}^{n,k},$$

we have following convergences

$$1. \sup_{|v| \leq b} \left| \mathbf{E} e^{v\Delta_{n,k}^*(\vartheta_0) - \frac{1}{2}v^2\mathbf{I}_{n,k}(\vartheta_0)} - 1 \right| \rightarrow 0.$$

$$2. \sup_{|\xi| \leq B} \left| \int_{\Omega} \xi d\mathbf{P}_{\vartheta_0 + \sqrt{\delta}v}^{n,k} - \int_{\Omega} \xi d\mathbf{Q}_{\vartheta_0, v}^{n,k} \right| \rightarrow 0,$$

for all  $B > 0$ ,  $b > 0$ , as  $n \rightarrow +\infty$ .

*Proof.* Since  $\Delta_{n,k}(\vartheta_0) \Rightarrow \Delta_t(\vartheta_0)$ , then, for each  $\alpha > 0$ ,  $\Delta_{n,k}^\alpha(\vartheta_0) \Rightarrow \Delta_t^\alpha(\vartheta_0)$ , where we have denoted

$$\Delta_{n,k}^\alpha(\vartheta_0) = \Delta_{n,k}(\vartheta_0) \mathbf{I}\{|\Delta_{n,k}(\vartheta_0)| \leq \alpha\},$$

$$\Delta_t^\alpha(\vartheta_0) = \Delta_t(\vartheta_0) \mathbf{I}\{|\Delta_t(\vartheta_0)| \leq \alpha\}.$$

From this weak convergence we derive

$$\sup_{|v| \leq \alpha} \left| \mathbf{E} e^{v\Delta_{n,k}^\alpha(\vartheta_0) - \frac{1}{2}v^2\mathbf{I}_{n,k}(\vartheta_0)} - \mathbf{E} e^{v\Delta_t^\alpha(\vartheta_0) - \frac{1}{2}v^2\mathbf{I}_t(\vartheta_0)} \right| \rightarrow 0,$$

as  $n \rightarrow +\infty$ , for each fixed  $\alpha > 0$ . This result is similar to the fact that from weak convergence (convergence of distribution functions) follows convergence of characteristic functions. Therefore, for all  $\alpha > 0$ ,  $\varepsilon > 0$ , there exists  $n_\varepsilon \in \mathcal{N}$  such that starting from number  $n_\varepsilon$  we have

$$\sup_{|v| \leq \alpha} \left| \mathbf{E} e^{v \Delta_{n,k}^\alpha(\vartheta_0) - \frac{1}{2} v^2 \mathbf{I}_{n,k}(\vartheta_0)} - \mathbf{E} e^{v \Delta_t^\alpha(\vartheta_0) - \frac{1}{2} v^2 \mathbf{I}_t(\vartheta_0)} \right| < \varepsilon,$$

hence there is a sub-sequence of natural numbers  $k_m \in \mathcal{N}$  such that for all  $n > k_m$ ,

$$\sup_{|v| \leq m} \left| \mathbf{E} e^{v \Delta_{n,k}^m(\vartheta_0) - \frac{1}{2} v^2 \mathbf{I}_{n,k}(\vartheta_0)} - \mathbf{E} e^{v \Delta_t^m(\vartheta_0) - \frac{1}{2} v^2 \mathbf{I}_t(\vartheta_0)} \right| < \frac{1}{m}.$$

Since one can always construct the sequence  $k_m$  in such a way that  $k_1 < k_2 < \dots < k_n < \dots$ , then, choosing a sequence  $m_n$  of natural numbers in the following way

$$k_{m_n} \leq n < k_{m_n+1}$$

(each natural number  $n$  is in some interval  $[k_j, k_{j+1})$ , so we are constructing correspondence between  $n$  and the index of the left point of that interval), we have that  $m_n$  is increasing, though not strictly,  $m_n \rightarrow +\infty$ , as  $n \rightarrow +\infty$ , and

$$\sup_{|v| \leq m_n} \left| \mathbf{E} e^{v \Delta_{n,k}^*(\vartheta_0) - \frac{1}{2} v^2 \mathbf{I}_{n,k}(\vartheta_0)} - \mathbf{E} e^{v \Delta_t^*(\vartheta_0) - \frac{1}{2} v^2 \mathbf{I}_t(\vartheta_0)} \right| < \frac{1}{m_n},$$

where  $\Delta_{n,k}^*(\vartheta_0) = \Delta_{n,k}^{m_n}(\vartheta_0)$  and  $\Delta_t^*(\vartheta_0) = \Delta_t^{m_n}(\vartheta_0)$ . Hence, for all  $b > 0$  we can write

$$\sup_{|v| \leq b} \left| \mathbf{E} e^{v \Delta_{n,k}^*(\vartheta_0) - \frac{1}{2} v^2 \mathbf{I}_{n,k}(\vartheta_0)} - \mathbf{E} e^{v \Delta_t^*(\vartheta_0) - \frac{1}{2} v^2 \mathbf{I}_t(\vartheta_0)} \right| \rightarrow 0,$$

as  $n \rightarrow +\infty$ . Now, we are going to show that for all  $b > 0$

$$\sup_{|v| \leq b} \left| \mathbf{E} e^{v \Delta_t^*(\vartheta_0) - \frac{1}{2} v^2 \mathbf{I}_t(\vartheta_0)} - 1 \right| \rightarrow 0,$$

as  $n$  tends to infinity. Using Lemma 2.1 we can calculate

$$\begin{aligned} & \sup_{|v| \leq b} \left| \mathbf{E} e^{v \Delta_t^*(\vartheta_0) - \frac{1}{2} v^2 \mathbf{I}_t(\vartheta_0)} - 1 \right| = \\ & \sup_{|v| \leq b} \left| \mathbf{E} e^{v \Delta_t^*(\vartheta_0) - \frac{1}{2} v^2 \mathbf{I}_t(\vartheta_0)} - \mathbf{E} e^{v \Delta_t(\vartheta_0) - \frac{1}{2} v^2 \mathbf{I}_t(\vartheta_0)} \right| = \\ & = \sup_{|v| \leq b} \int_{\{|\Delta_t(\vartheta_0)| > m_n\}} e^{-\frac{1}{2} v^2 \mathbf{I}_t(\vartheta_0)} d\mathbf{P} \leq \\ & \int_{\{|\Delta_t(\vartheta_0)| > m_n\}} d\mathbf{P} = \mathbf{P}\{|\Delta_t(\vartheta_0)| > m_n\} \rightarrow 0, \end{aligned}$$



as  $n \rightarrow +\infty$ , since  $I_t(\vartheta_0) > 0$  and  $m_n \rightarrow +\infty$ , as  $n \rightarrow +\infty$ . Combining two previous convergences we get

$$\begin{aligned} & \sup_{|v| \leq b} \left| \mathbf{E} e^{v\Delta_{n,k}^*(\vartheta_0) - \frac{1}{2}v^2 I_t(\vartheta_0)} - 1 \right| \leq \\ & \sup_{|v| \leq b} \left| \mathbf{E} e^{v\Delta_{n,k}^*(\vartheta_0) - \frac{1}{2}v^2 I_t(\vartheta_0)} - \mathbf{E} e^{v\Delta_t^*(\vartheta_0) - \frac{1}{2}v^2 I_t(\vartheta_0)} \right| + \\ & + \sup_{|v| \leq b} \left| \mathbf{E} e^{v\Delta_t^*(\vartheta_0) - \frac{1}{2}v^2 I_t(\vartheta_0)} - 1 \right| \rightarrow 0, \quad n \rightarrow +\infty, \end{aligned}$$

Hence,

$$\sup_{|v| \leq b} \left| \mathbf{E} e^{v\Delta_{n,k}^*(\vartheta_0) - \frac{1}{2}v^2 I_t(\vartheta_0)} - 1 \right| \rightarrow 0.$$

Once more using Lemma 2.1, we can show that

$$\mathbf{Z}_{n,k}(v) = \frac{d\mathbf{P}_{\vartheta_0 + \sqrt{\delta}v}^{n,k}}{d\mathbf{P}_{\vartheta_0}^{n,k}}(\mathbf{X}^k), \quad \hat{\mathbf{Z}}_{n,k}(v) = c_{n,k}(\vartheta_0, v) e^{v\Delta_{n,k}^*(\vartheta_0) - \frac{1}{2}v^2 I_t(\vartheta_0)} \quad (2.15)$$

are uniformly integrable. Indeed from LAMN condition we have

$$\mathbf{Z}_{n,k}(v) \Longrightarrow \xi, \quad \mathbf{E}\xi = 1,$$

(in fact this means contiguity of corresponding measures), and using Lemma 2.1 we also obtain

$$\lim_{n \rightarrow +\infty} \mathbf{E}_{\vartheta_0} \mathbf{Z}_{n,k}(v) = 1,$$

that is, we can pass to the limit under the sign of the integral, which is possible if and only if the sequence  $\mathbf{Z}_{n,k}(v)$  is uniformly integrable. In the same way, since

$$\hat{\mathbf{Z}}_{n,k}(v) - \mathbf{Z}_{n,k}(v) \rightarrow 0,$$

in  $\mathbf{P}_{\vartheta_0}^{n,k}$  probability, as  $n \rightarrow +\infty$ , then

$$\hat{\mathbf{Z}}_{n,k}(v) = (\hat{\mathbf{Z}}_{n,k}(v) - \mathbf{Z}_{n,k}(v)) + \mathbf{Z}_{n,k}(v) \Longrightarrow \xi, \quad \mathbf{E}\xi = 1.$$

Finally, for all  $v \in R$

$$\lim_{n \rightarrow +\infty} \mathbf{E}_{\vartheta_0} \hat{\mathbf{Z}}_{n,k}(v) = 1,$$

that is,  $\hat{\mathbf{Z}}_{n,k}(v)$  is uniformly integrable too. Hence second assertion of the theorem follows.  $\blacksquare$

**Conditions  $\mathcal{R}$ .** Diffusion process (2.14) is such that

1. With respect to  $\vartheta$ , the diffusion coefficient  $\sigma(\vartheta, t, x)$  has the derivative and

$$|\dot{\sigma}(\vartheta, t, x) - \dot{\sigma}(\tilde{\vartheta}, t, x)| \leq H|\tilde{\vartheta} - \vartheta|^\alpha, \quad \forall \tilde{\vartheta}, \vartheta \in \Theta, \quad \alpha > 0,$$

2. With respect to  $x$ , the function  $\dot{\sigma}(\vartheta, t, x)$  is of class  $C^2$  (twice continuously differentiable), the functions  $\sigma(\vartheta, t, x), S(t, x)$  are of class  $C^3$ ,
3. With respect to  $t$ , the functions

$$S(t, x), \sigma(\vartheta, t, x), \dot{\sigma}(\vartheta, t, x), \sigma'(\vartheta, t, x), S'(t, x)$$

are of class  $C^1$ ,

4. For the diffusion and drift coefficients we have

$$|\sigma'(\vartheta, t, x)| \leq A, \quad |S'(t, x)| \leq A, \quad \sigma(\vartheta, t, x) > 0,$$

and

$$\left| \frac{\dot{\sigma}(\vartheta, t, x)}{\sigma(\vartheta, t, x)} \right| \geq \kappa > 0, \quad \forall (\vartheta, t, x) \in \Theta \times [0, T] \times \mathbf{R},$$

5.  $\dot{u}(\vartheta, t, x), u'(\vartheta, t, x), \dot{u}'(\vartheta, t, x)$  are continuous w.r.t. the parameter  $\vartheta$  and

$$|\phi(t, x, \vartheta)| \leq C(1 + |x|^p), \quad p > 0,$$

where  $\phi(t, x, \vartheta) = \dot{u}(\vartheta, t, x), u'(\vartheta, t, x), \dot{u}'(\vartheta, t, x)$ .

**Theorem 2.3.** *Suppose that the coefficients of the diffusion process (2.14) satisfies  $\mathcal{R}$  conditions, then, for the loss function  $\ell(u) = |u|^p, p > 0$ , the following lower bound is true*

$$\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow +\infty} \sup_{|\vartheta - \vartheta_0| < \varepsilon} \mathbf{E}_{\vartheta} \ell(\delta^{-1/2} (\bar{Y}_{t_k, n} - Y_{t_k})) \geq \mathbf{E}_{\vartheta_0} \ell(\dot{u}(\vartheta_0, t, X_t) \xi_t(\vartheta_0)),$$

$$\xi_t(\vartheta_0) = \frac{\Delta_t(\vartheta_0)}{I_t(\vartheta_0)}.$$

*Proof.* We are following the ideas of the proof of [14]. Denote

$$A = \sup_{|\vartheta - \vartheta_0| < \varepsilon} \mathbf{E}_{\vartheta} \ell(\delta^{-1/2} (\bar{Y}_{t_k, n} - Y_{t_k})) \geq$$

$$\geq \frac{1}{2\varepsilon} \int_{|\vartheta - \vartheta_0| < \varepsilon} \mathbf{E}_{\vartheta} \{ \ell[\delta^{-1/2} (\bar{Y}_{t_k, n} - Y_{t_k})] \} d\vartheta,$$

Using the change of variable  $\vartheta = \vartheta_0 + \sqrt{\delta}v$  one has

$$A \geq \frac{1}{2b} \int_{|v| < b} \mathbf{E}_{\vartheta_0 + \sqrt{\delta}v} \{ \ell[\delta^{-1/2} (\bar{Y}_{t_k, n} - Y_{t_k})] \} dv,$$

for any constants  $b > 0$ ,  $\varepsilon > 0$ , as  $n$  is sufficiently large. Using the Taylor's expansion we obtain

$$\begin{aligned} Y_{t_k} &= u(t_k, X_{t_k}, \vartheta) = u(t_k, X_{t_k}, \vartheta_0 + \sqrt{\delta}v) = \\ &= u(t_k, X_{t_k}, \vartheta_0) + \sqrt{\delta}v\dot{u}(t_k, X_{t_k}, \vartheta_0) + \alpha(v)\sqrt{\delta}v = \\ &= Y_{t_k}^0 + \dot{u}(t_k, X_{t_k}, \tilde{\vartheta})\sqrt{\delta}v, \quad 0 < \alpha(v) < 1. \end{aligned}$$

Here we have denoted

$$Y_{t_k}^0 = u(t_k, X_{t_k}, \vartheta_0), \quad \tilde{\vartheta} = \vartheta_0 + \alpha(v)\sqrt{\delta}v.$$

Therefore we obtain the inequality

$$A \geq \frac{1}{2b} \int_{|v|<b} \mathbf{E}_{\vartheta_0 + \sqrt{\delta}v} \left\{ \ell \left[ \delta^{-1/2} (\bar{Y}_{t_k, n} - Y_{t_k}^0) - v\dot{u}(t_k, X_{t_k}, \tilde{\vartheta}) \right] \right\} dv.$$

Denote  $\ell_a(x) = \min(a, \ell(x))$ . Since the loss function is positive, then, changing the measure, one has

$$\begin{aligned} A &\geq \frac{1}{2b} \int_{|v|<b} \mathbf{E}_{\vartheta_0} \left\{ \ell_a \left[ \delta^{-1/2} (\bar{Y}_{t_k, n} - Y_{t_k}^0) - v\dot{u}(t_k, X_{t_k}, \tilde{\vartheta}) \right] \frac{d\mathbf{P}_{\vartheta_0 + \sqrt{\delta}v}^n(\mathbf{X}^k)}{d\mathbf{P}_{\vartheta_0}^n(\mathbf{X}^k)} \right\} dv \\ &= \frac{1}{2b} \int_{|v|<b} \mathbf{E}_{\vartheta_0} \left\{ \ell_a \left[ \delta^{-1/2} (\bar{Y}_{t_k, n} - Y_{t_k}^0) - v\dot{u}(t_k, X_{t_k}, \tilde{\vartheta}) \right] Z_{n, k}(v) \right\} dv \\ &= \frac{1}{2b} \int_{|v|<b} \mathbf{E}_{\vartheta_0} \left\{ \ell_a \left[ \delta^{-1/2} (\bar{Y}_{t_k, n} - Y_{t_k}^0) - v\dot{u}(t_k, X_{t_k}, \tilde{\vartheta}) \right] \left( Z_{n, k}(v) - \hat{Z}_{n, k}(v) \right) \right\} dv \\ &\quad + \frac{1}{2b} \int_{|v|<b} \mathbf{E}_{\vartheta_0} \left\{ \ell_a \left[ \delta^{-1/2} (\bar{Y}_{t_k, n} - Y_{t_k}^0) - v\dot{u}(t_k, X_{t_k}, \tilde{\vartheta}) \right] \hat{Z}_{n, k}(v) \right\} dv \\ &= \frac{1}{2b} \int_{|v|<b} \mathbf{E}_{\vartheta_0} \left\{ \ell_a \left[ \delta^{-1/2} (\bar{Y}_{t_k, n} - Y_{t_k}^0) - v\dot{u}(t_k, X_{t_k}, \tilde{\vartheta}) \right] \hat{Z}_{n, k}(v) \right\} dv + o(1). \end{aligned}$$

Here we have used the Proposition 2.3. Suppose that the Lipschitz condition holds for the derivative

$$|\dot{u}(t_k, X_{t_k}, \vartheta_0 + \alpha(v)\sqrt{\delta}v) - \dot{u}(t_k, X_{t_k}, \vartheta_0)| \leq L|\alpha(v)|\sqrt{\delta}|v|,$$

since  $\ell_a(u)$  also satisfies the Lipschitz condition

$$|\ell_a(u) - \ell_a(v)| \leq L_0|u - v|,$$

then,

$$\begin{aligned} &\left| \ell_a \left[ \delta^{-1/2} (\bar{Y}_{t_k, n} - Y_{t_k}^0) - v\dot{u}(t_k, X_{t_k}, \tilde{\vartheta}) \right] - \ell_a \left[ \delta^{-1/2} (\bar{Y}_{t_k, n} - Y_{t_k}^0) - v\dot{u}(t_k, X_{t_k}, \vartheta_0) \right] \right| \\ &\leq L_0|v| |\dot{u}(t_k, X_{t_k}, \tilde{\vartheta}) - \dot{u}(t_k, X_{t_k}, \vartheta_0)| \leq L_0L|v|^2|\alpha(v)|\sqrt{\delta} \leq L_0Lb^2\sqrt{\delta}, \end{aligned}$$

Further,

$$e^{v\hat{\Delta}_{n,k}(\vartheta_0) - \frac{1}{2}|v|^2 I_{n,k}(\vartheta_0)} \leq e^{v\hat{\Delta}_{n,k}(\vartheta_0)} \leq C,$$

since  $|v| < b$  and  $\hat{\Delta}_{n,k}(\vartheta_0)$  is a truncation of some vector (hence, it is bounded). Then,

$$\frac{1}{2b} \int_{|v|<b} \mathbf{E}_{\vartheta_0} \left\{ L_0 L b^2 \sqrt{\delta} e^{v\hat{\Delta}_{n,k}(\vartheta_0) - \frac{1}{2}|v|^2 I_{n,k}(\vartheta_0)} \right\} dv \leq C\sqrt{\delta},$$

where  $C$  is a constant. Hence, this term is  $o(1)$  as  $\delta \rightarrow 0$ . Finally we have such an inequality

$$A \geq \frac{1}{2b} \int_{|v|<b} \mathbf{E}_{\vartheta_0} \left\{ \ell_a \left[ \delta^{-1/2} (\bar{Y}_{t_k,n} - Y_{t_k}^0) - v \dot{u}(t_k, X_{t_k}, \vartheta_0) \right] \hat{Z}_{n,k}(v) \right\} dv + o(1).$$

Denote

$$\xi_n = e^{\frac{1}{2} \frac{\hat{\Delta}_{n,k}(\vartheta_0)^2}{I_{n,k}(\vartheta_0)}}, \quad y = v \sqrt{I_{n,k}(\vartheta_0)} - \frac{\hat{\Delta}_{n,k}(\vartheta_0)}{\sqrt{I_{n,k}(\vartheta_0)}}$$

$$\eta_n = \frac{\delta^{-1/2} (\bar{Y}_{t_k,n} - Y_{t_k}^0)}{\dot{u}(t_k, X_{t_k}, \vartheta_0)} + \frac{\hat{\Delta}_{n,k}(\vartheta_0)}{I_{n,k}(\vartheta_0)}.$$

Using these notations and changing the order of integration by Fubini's theorem one can obtain

$$A \geq \frac{1}{2b} \mathbf{E}_{\vartheta_0} \left\{ \xi_n \int_{\tilde{K}_b} \ell_a \left[ \dot{u}(t_k, X_{t_k}, \vartheta_0) \left( \eta_n - \frac{y}{\sqrt{I_{n,k}(\vartheta_0)}} \right) \right] e^{-\frac{1}{2}|y|^2} dy \right\} + o(1),$$

where

$$\tilde{K}_b = \left\{ \left| \frac{y}{\sqrt{I_{n,k}(\vartheta_0)}} + \frac{\hat{\Delta}_{n,k}(\vartheta_0)}{I_{n,k}(\vartheta_0)} \right| < b \right\}.$$

Denote

$$C = \left\{ \left| \frac{\hat{\Delta}_{n,k}(\vartheta_0)}{I_{n,k}(\vartheta_0)} \right| < b - \frac{\sqrt{b}}{\sqrt{I_{n,k}(\vartheta_0)}} \right\}.$$

Then, on this set  $K_{\sqrt{b}} = \{|y| < \sqrt{b}\} \subset \tilde{K}_b$ . Since  $\xi_n > 0$  and the loss function is positive,

$$A \geq \frac{1}{2b} \int_C \xi_n \int_{K_{\sqrt{b}}} \ell_a \left[ \dot{u}(t_k, X_{t_k}, \vartheta_0) \left( \eta_n - \frac{y}{\sqrt{I_{n,k}(\vartheta_0)}} \right) \right] e^{-\frac{1}{2}|y|^2} dy d\mathbf{P}_{\vartheta_0}^n + o(1).$$

By the Anderson's lemma one has

$$\begin{aligned} & \int_{K_{\sqrt{b}}} \ell_a \left[ \dot{u}(t_k, X_{t_k}, \vartheta_0) \left( \eta_n - \frac{y}{\sqrt{I_{n,k}(\vartheta_0)}} \right) \right] e^{-\frac{1}{2}|y|^2} dy \geq \\ & \geq \int_{K_{\sqrt{b}}} \ell_a \left[ \dot{u}(t_k, X_{t_k}, \vartheta_0) \frac{y}{\sqrt{I_{n,k}(\vartheta_0)}} \right] e^{-\frac{1}{2}|y|^2} dy = J(a, b). \end{aligned}$$

Therefore,

$$\begin{aligned} A & \geq \frac{1}{2b} \int_C \xi_n \int_{K_{\sqrt{b}}} \ell_a \left[ \dot{u}(t_k, X_{t_k}, \vartheta_0) \frac{y}{\sqrt{I_{n,k}(\vartheta_0)}} \right] e^{-\frac{1}{2}|y|^2} dy d\mathbf{P}_{\vartheta_0}^n \\ & + o(1) = \frac{J(a, b)}{2b} \int_C \xi_n d\mathbf{P}_{\vartheta}^{n,k} + o(1) \\ & = \frac{J(a, b)}{2b} \mathbf{E}_{\vartheta}^{n,k} \left\{ \xi_n \mathbb{I} \left\{ \left| \frac{\hat{\Delta}_{n,k}(\vartheta_0)}{I_{n,k}(\vartheta_0)} \right| < b - \frac{\sqrt{b}}{\sqrt{I_{n,k}(\vartheta_0)}} \right\} \right\} + o(1). \end{aligned}$$

Passing to the limit we get

$$\lim_{n \rightarrow +\infty} A \geq \frac{J(a, b)}{\sqrt{2\pi}} \frac{2b - \sqrt{b}}{2b}.$$

Since  $a, b$  are arbitrary, then

$$\lim_{\substack{a \rightarrow +\infty \\ b \rightarrow +\infty}} J(a, b) = \mathbf{E}_{\vartheta_0} \ell(\dot{u}(\vartheta_0, t, X_t) \xi_t(\vartheta_0)).$$

■

### 2.5.5 Asymptotically efficient approximation of the solution of a FBSDE

Introduce for the values  $t_k \in [\tau, T]$  the estimators

$$Y_{t_k, n}^* = u(t_k, X_{t_k}, \vartheta_{k, n}^*), \quad Z_{t_k, n}^* = u'_x(t_k, X_{t_k}, \vartheta_{k, n}^*) \sigma(t_k, X_{t_k}, \vartheta_{k, n}^*).$$

**Theorem 2.4.** *Suppose that the conditions of regularity hold, then the estimators*

$$(Y_{t, n}^*, t \in [\tau, T]) \quad \text{and} \quad (Z_{t, n}^*, t \in [\tau, T])$$

are consistent

$$Y_{t_k,n}^* \longrightarrow Y_t, \quad Z_{t_k,n}^* \longrightarrow Z_t,$$

and are asymptotically conditionally normal (stable convergence)

$$\begin{aligned} \delta^{-1/2} (Y_{t_k,n}^* - Y_{t_k}) &\Longrightarrow \langle \dot{u}(t, X_t, \vartheta_0), \xi_t(\vartheta_0) \rangle, \\ \delta^{-1/2} (Z_{t_k,n}^* - Z_{t_k}) &\Longrightarrow \sigma(t, X_t, \vartheta_0) \langle \dot{u}'_x(t, X_t, \vartheta_0), \xi_t(\vartheta_0) \rangle \\ &\quad + u'_x(t, X_t, \vartheta_0) \langle \dot{\sigma}(t, X_t, \vartheta_0), \xi_t(\vartheta_0) \rangle. \end{aligned}$$

*Proof.* Let us denote  $v_{k,n}^* = \delta^{-1/2} (\vartheta_{k,n}^* - \vartheta_0)$  and write the Taylor expansion

$$\begin{aligned} Y_{t_k,n}^* &= u(t_k, X_{t_k}, \vartheta_0 + \delta^{1/2} v_{k,n}^*) = u(t_k, X_{t_k}, \vartheta_0) \\ &\quad + \delta^{1/2} \langle v_{k,n}^*, \dot{u}(t_k, X_{t_k}, \vartheta_0) \rangle + o(\delta^{-1/2}), \\ Z_{t_k,n}^* &= u'_x(t_k, X_{t_k}, \vartheta_0 + \delta^{1/2} v_{k,n}^*) \sigma(t_k, X_{t_k}, \vartheta_0 + \delta^{1/2} v_{k,n}^*) \\ &= \sigma(t_k, X_{t_k}, \vartheta_0) \langle \dot{u}'_x(t_k, X_{t_k}, \vartheta_0), \xi_{t_k}(\vartheta_0) \rangle \\ &\quad + u'_x(t_k, X_{t_k}, \vartheta_0) \langle \dot{\sigma}(t_k, X_{t_k}, \vartheta_0), \xi_{t_k}(\vartheta_0) \rangle + o(\delta^{-1/2}). \end{aligned}$$

Now the proof follows from the Theorem 2.2 and the regularity of the functions  $u(\cdot)$ ,  $u'_x(\cdot)$  and  $\sigma(\cdot)$ .  $\blacksquare$

**Remark 2.4.** Note that we do not evaluate the difference  $\delta^{-1/2}(Y_{t,n}^* - Y_t)$  for  $t \in [t_k, t_{k+1})$  because in the representation

$$\delta^{-1/2}(Y_{t,n}^* - Y_t) = \delta^{-1/2}(Y_{t,n}^* - Y_{t_k}) + \delta^{-1/2}(Y_{t_k} - Y_t)$$

for the second term we have the relation

$$\begin{aligned} \delta^{-1/2}(Y_t - Y_{t_k}) &= \delta^{-1/2} u'_x(t, \tilde{X}_{t_k}, \vartheta_0) (X_t - X_{t_k}) \\ &= u'_x(t, X_{t_k}, \vartheta_0) \sigma(\vartheta_0, t_k, X_{t_k}) \frac{(W_t - W_{t_k})}{\sqrt{t_{k+1} - t_k}} (1 + o(1)) \\ &= u'_x(t, X_{t_k}, \vartheta_0) \sigma(\vartheta_0, t_k, X_{t_k}) \zeta_t \sqrt{v_n(t)} (1 + o(1)) \end{aligned}$$

Here  $\zeta_t \sim \mathcal{N}(0, 1)$  and  $v_n(t) = (t_{k+1} - t_k)^{-1} (t - t_k)$ . The study of the limit of  $v_n(t)$  for all  $t \in [0, T]$  is a special interesting problem.  $\diamond$

To prove the optimality of the presented approximations  $Y_{t_k,n}^*$  and  $Z_{t_k,n}^*$  we need the lower bound of Jeganathan type given in the following proposition. Below

$$\eta(t, X_t, \vartheta_0) = \langle \dot{u}'_x(t, X_t, \vartheta_0), \xi_t(\vartheta_0) \rangle + u'_x(t, X_t, \vartheta_0) \langle \dot{\sigma}(t, X_t, \vartheta_0), \xi_t(\vartheta_0) \rangle$$

**Proposition 2.4.** *Let the conditions of regularity be fulfilled. Then for all estimators  $\bar{Y}_{t_k, n}$  and  $\bar{Z}_{t_k, n}$  we have*

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \sup_{|\vartheta - \vartheta_0| \leq \varepsilon} \mathbf{E}_{\vartheta} \ell \left( \delta^{-1/2} (\bar{Y}_{t_k, n} - Y_{t_k}) \right) &\geq \mathbf{E}_{\vartheta_0} \ell \left( \langle \dot{u}(t, X_t, \vartheta_0), \xi_t(\vartheta_0) \rangle \right), \\ \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \sup_{|\vartheta - \vartheta_0| \leq \varepsilon} \mathbf{E}_{\vartheta} \ell \left( \delta^{-1/2} (\bar{Z}_{t_k, n} - Z_{t_k}) \right) &\geq \mathbf{E}_{\vartheta_0} \ell \left( \eta(t, X_t, \vartheta_0) \right). \end{aligned}$$

The proof of this proposition is similar to the proof of the lower bound in the problem of approximation of the solution of BSDE in the asymptotics of small noise given in the works [24].

**Example.** Black-Scholes model. We are given the forward equation

$$dX_t = \alpha X_t dt + \vartheta X_t dW_t, \quad X_0 = x_0, \quad 0 \leq t \leq T$$

and two functions  $f(x, y, z) = \beta y + \gamma xz$  and  $\Phi(x)$ . We have to approximate the solution of the backward equation

$$dY_t = -\beta Y_t dt - \gamma X_t Z_t dt + Z_t dW_t, \quad Y_T = \Phi(X_T)$$

in the situation where  $\vartheta \in (a, b)$ ,  $a > 0$  and is unknown.

The corresponding partial differential equation is

$$\frac{\partial u}{\partial t} + (\alpha + \vartheta \gamma) x \frac{\partial u}{\partial x} + \frac{\vartheta^2 x^2}{2} \frac{\partial^2 u}{\partial x^2} + \beta u = 0, \quad u(T, x, \vartheta) = \Phi(x).$$

The solution of this equation is the function

$$u(t, x, \vartheta) = \frac{e^{\beta(T-t)}}{\sqrt{2\pi\vartheta^2(T-t)}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2\vartheta^2(T-t)}} \Phi \left( x e^{(\alpha + \vartheta \gamma - \frac{\vartheta^2}{2})(T-t) - z} \right) dz.$$

The discrete-time observations are  $X^n = (X_0, X_{t_1}, \dots, X_{t_n})$ . We can calculate the pseudo MLE-process

$$\hat{\vartheta}_{t_k, n} = \left( \frac{1}{t_k} \sum_{j=0}^{k-1} \frac{(X_{t_{j+1}} - X_{t_j} - \alpha X_{t_j} \delta)^2}{X_{t_j}^2} \right)^{\frac{1}{2}}.$$

The estimator of  $Y_t = u(t, X_t, \vartheta_0)$  is

$$\hat{Y}_{t_k} = \int_{-\infty}^{\infty} \frac{e^{-\frac{z^2}{2\hat{\vartheta}_{t_k, n}^2(T-t_k)} + \beta(T-t_k)}}{\sqrt{2\pi\hat{\vartheta}_{t_k, n}^2(T-t_k)}} \Phi \left( X_{t_k} e^{(\alpha + \hat{\vartheta}_{t_k, n} \gamma - \frac{\hat{\vartheta}_{t_k, n}^2}{2})(T-t_k) - z} \right) dz,$$

where  $k = \lceil \frac{t}{T}n \rceil$ .

Approximation of  $\hat{Z}_t = \vartheta_0 X_t u'(t, X_t, \vartheta_0)$  can be written explicitly as well.

The one-step MLE-process is constructed as follows. Let us fix a learning interval  $[0, \tau]$ ,  $\tau < T$  and introduce the estimator  $\hat{\vartheta}_{\tau, n}$  constructed by the observations  $X^N = (X_0, X_{t_1}, \dots, X_{t_N})$ , where  $N = \lceil \frac{n\tau}{T} \rceil$  as preliminary. Then we have

$$\vartheta_{k, n}^* = \hat{\vartheta}_{\tau, n} + \frac{1}{2\hat{\vartheta}_{\tau, n}} \sum_{j=1}^k \left[ (X_j - X_{t_{j-1}} - \alpha X_{t_{j-1}} \delta)^2 - \hat{\vartheta}_{\tau, n}^2 X_{t_{j-1}}^2 \delta \right].$$

The corresponding approximations are

$$\hat{Y}_{t_k}^* = u(t_k, X_{t_k}, \vartheta_{k, n}^*), \quad Z_t^* = \vartheta_{k, n}^* X_t u'(t, X_t, \vartheta_{k, n}^*), \quad N+1 \leq k \leq n$$

and by Theorem 2.4 and by Proposition 2.4 these approximations are asymptotically conditionally normal and asymptotically efficient.

## 2.5.6 Pearson diffusion

To illustrate the results obtained in the first chapter we consider an example of a parameter estimation problem for the particular model of observations of Pearson-type diffusion process

$$dX_t = -X_t dt + \sqrt{\vartheta + X_t^2} dW_t, \quad X_0, \quad 0 \leq t \leq T. \quad (2.16)$$

Here  $\vartheta \in \Theta = (\alpha, \beta)$ ,  $\alpha > 0$  is unknown parameter.

Note that this is a particular case of the family of stochastic processes known as *Pearson diffusions* [38], section 1.3.7.

It is easy to see that in the case of continuous time observation the problem of parameter estimation is degenerated (singular), i.e. the unknown parameter  $\vartheta$  can be estimated without error. Indeed, by Itô's formula we can write

$$X_t^2 = X_0^2 + 2 \int_0^t X_s dX_s + \int_0^t [\vartheta + X_s^2] ds.$$

Hence for all  $t \in (0, T]$  we have the equality

$$\hat{\vartheta} = t^{-1} \left[ X_t^2 - X_0^2 - 2 \int_0^t X_s dX_s - \int_0^t X_s^2 ds \right] \quad (2.17)$$



and  $\hat{\vartheta} = \vartheta$ . This effect is due to the singularity of the measures induced by the observations in the space of realizations.

Such problems of parameter estimation in the diffusion coefficient are usually studied in the case of discrete time observations  $X^n = (X_{t_0}, X_{t_1}, \dots, X_{t_n})$ , where  $0 = t_0 < t_1 < \dots < t_n = T$ . Then the problem is no more singular and became an interesting statistical estimation problem. There is a diversity of the choice of observing times  $t_i$ . In this section we present results from the paper [7]. We take the simplest case of equidistant observations, i.e.  $t_j = j\delta, \delta = \frac{T}{n}$  and we study the properties of the estimators in the asymptotics of *high frequency* as  $n \rightarrow \infty$ . Our goal is to construct an asymptotically efficient estimator of the parameter  $\vartheta$ . Note that the family of measures induced by the observations  $X^k = (X_{t_0}, X_{t_1}, \dots, X_{t_k})$  with  $t_k$  satisfying  $t_k \leq t < t_{k+1}$  and fixed  $t$  are *locally asymptotically mixed normal* (LAMN) and for all estimators  $\vartheta_k^*$  we have the lower bound for the risk

$$\lim_{\nu \rightarrow 0} \lim_{n \rightarrow \infty} \sup_{|\vartheta - \vartheta_0| < \nu} \mathbf{E}_{\vartheta} \ell \left( \sqrt{k} (\vartheta_k^* - \vartheta) \right) \geq \mathbf{E}_{\vartheta_0} \ell (\zeta_t (\vartheta_0)). \quad (2.18)$$

Here as a loss function  $\ell(\cdot)$  can be taken, for example, polynomial  $\ell(u) = |u|^p, p > 0$ . For the definition of the random function  $\zeta_t(\vartheta_0)$  see (2.22) below. An estimator  $\vartheta_k^*$  is called *asymptotically efficient* if for all  $\vartheta_0 \in \Theta$  we have

$$\lim_{\nu \rightarrow 0} \lim_{n \rightarrow \infty} \sup_{|\vartheta - \vartheta_0| < \nu} \mathbf{E}_{\vartheta} \ell \left( \sqrt{k} (\vartheta_k^* - \vartheta) \right) = \mathbf{E}_{\vartheta_0} \ell (\zeta_t (\vartheta_0)). \quad (2.19)$$

The proof of this statement can be found in [4] and [16].

The estimator we construct in two steps. First we propose a consistent estimator  $\bar{\vartheta}_N$  of this parameter based on the first  $N$  observations  $X^N = (X_{t_0}, X_{t_1}, \dots, X_{t_N})$  on the time interval  $[0, \tau]$ . Here  $\tau = t_N = N\frac{T}{n}$ . Then using this estimator and one-step type device we propose an asymptotically efficient estimator.

The first consistent estimator we obtain from the equality (2.17) by replacing the integrals by the corresponding integral sums

$$\bar{\vartheta}_N = \frac{n}{TN} \left[ X_{t_N}^2 - X_0^2 - 2 \sum_{j=1}^N X_{t_{j-1}} [X_{t_j} - X_{t_{j-1}}] - \sum_{j=1}^N X_{t_{j-1}}^2 \delta \right].$$

The consistency of this estimator follows immediately from the limits

$$\sum_{j=1}^N X_{t_{j-1}} [X_{t_j} - X_{t_{j-1}}] \longrightarrow \int_0^\tau X_s dX_s, \quad \sum_{j=1}^N X_{t_{j-1}}^2 \delta \longrightarrow \int_0^\tau X_s^2 ds$$

and the relation (2.17).

The next step is to see the behavior of the error of estimation.

Consider  $\xi_N = \sqrt{N} (\bar{\vartheta}_N - \vartheta)$ . Then

$$\xi_N = \frac{\sqrt{N}}{\tau} \left[ \int_0^\tau X_s [2dX_s + X_s ds] - \sum_{j=1}^N X_{t_{j-1}} [2(X_{t_j} - X_{t_{j-1}}) + X_{t_{j-1}} \delta] \right].$$

We have

$$\begin{aligned} X_{t_{j-1}} [X_{t_j} - X_{t_{j-1}}] - \int_{t_{j-1}}^{t_j} X_s dX_s &= \int_{t_{j-1}}^{t_j} [X_{t_{j-1}} - X_s] dX_s \\ &= - \int_{t_{j-1}}^{t_j} X_s [X_{t_{j-1}} - X_s] ds + \int_{t_{j-1}}^{t_j} [X_{t_{j-1}} - X_s] \sqrt{\vartheta_0 + X_s^2} dW_s \\ &= O(\delta^{3/2}) + \int_{t_{j-1}}^{t_j} \left( \int_{t_{j-1}}^s \sqrt{\vartheta_0 + X_r^2} dW_r \right) \sqrt{\vartheta_0 + X_s^2} dW_s \\ &= O(\delta^{3/2}) + (\vartheta_0 + X_{t_{j-1}}^2) \left[ \frac{(W_{t_j} - W_{t_{j-1}})^2 - \delta}{2} \right] \\ &= O(\delta^{3/2}) + (\vartheta_0 + X_{t_{j-1}}^2) \sqrt{\frac{\delta}{2}} w_j, \end{aligned}$$

where we used the estimate  $X_{t_j} - X_{t_{j-1}} = O(\delta^{1/2})$  and denoted

$$w_j = \frac{(W_{t_j} - W_{t_{j-1}})^2 - \delta}{\sqrt{2\delta}}, \quad \mathbf{E}w_j = 0, \quad \mathbf{E}w_j^2 = \delta, \quad \mathbf{E}w_j w_i = 0, \quad i \neq j.$$

We have as well

$$X_{t_{j-1}}^2 \delta - \int_{t_{j-1}}^{t_j} X_s^2 ds = \int_{t_{j-1}}^{t_j} [X_{t_{j-1}}^2 - X_s^2] ds = O(\delta^{3/2}).$$

Therefore we obtain the stable convergence (see [38])

$$\xi_N = \frac{\sqrt{2\delta N}}{\tau} \sum_{j=1}^N (\vartheta_0 + X_{t_{j-1}}^2) w_j + o(1) \implies \xi_\tau = \sqrt{\frac{2}{\tau}} \int_0^\tau (\vartheta_0 + X_s^2) dw(s).$$

More detailed analysis shows that we have the convergence of moments too: for any  $p > 0$

$$n^{\frac{p}{2}} \mathbf{E}_{\vartheta_0} |\bar{\vartheta}_N - \vartheta_0|^p \longrightarrow \mathbf{E}_{\vartheta_0} |\xi_\tau|^p.$$

The pseudo log-likelihood function is

$$L(\vartheta, X^N) = -\frac{1}{2} \sum_{j=1}^N \ln \left( 2\pi \left( \vartheta + X_{t_{j-1}}^2 \right) \right) - \sum_{j=1}^N \frac{[X_{t_j} - X_{t_{j-1}} + X_{t_{j-1}} \delta]^2}{2 \left( \vartheta + X_{t_{j-1}}^2 \right) \delta}.$$

It is easy to see that the equation  $\dot{L}(\vartheta, X^N) = 0$  has no solution, which can be written in explicit form. Hence the corresponding pseudo MLE cannot be written in explicit form too. Remind that this estimator is asymptotically efficient [4], [9].

Our goal is to use the well-known one-step MLE device [25], [26] in the construction of one-step MLE-process. This estimator-process is asymptotically equivalent to the pseudo MLE but can be calculated in explicit form. This type of estimator-processes were proposed in [23] in the problem of approximation of the solution of backward stochastic differential equation for several models of observations (see the review of recent results in [22]).

Let us fix  $t \in (\tau, T]$  and take such  $k$  that  $t_k \leq t < t_{k+1}$ . As  $n \rightarrow \infty$  we have  $k \rightarrow \infty$  and  $t_k \rightarrow t$ . We consider the estimation of  $\vartheta$  by the observations  $X^k = (X_{t_0}, X_{t_1}, \dots, X_{t_k})$ . Recall that by the first  $X^N$  observations we already obtained the estimator  $\bar{\vartheta}_N$ . Denote the pseudo Fisher information

$$\mathbb{I}_{t_k, n}(\vartheta_0) = \frac{1}{2} \sum_{j=1}^k \frac{\delta}{\left( \vartheta_0 + X_{t_{j-1}}^2 \right)^2} \longrightarrow \mathbb{I}_t(\vartheta_0) = \frac{1}{2} \int_0^t \frac{ds}{\left( \vartheta_0 + X_s^2 \right)^2}, \quad n \longrightarrow +\infty.$$

The one-step MLE-process introduced first in [23] is

$$\vartheta_{t_k, n}^* = \bar{\vartheta}_N + \sqrt{\delta} \sum_{j=1}^k \frac{[X_{t_j} - X_{t_{j-1}} + X_{t_{j-1}} \delta]^2 - \left( \bar{\vartheta}_N + X_{t_{j-1}}^2 \right) \delta}{2 \mathbb{I}_{t_k, n}(\bar{\vartheta}_N) \left( \bar{\vartheta}_N + X_{t_{j-1}}^2 \right)^2 \sqrt{\delta}}, \quad \tau \leq t_k \leq T.$$

Let us denote

$$\Delta_{t_k, n}(\vartheta, X^k) = \sum_{j=1}^k \frac{[X_{t_j} - X_{t_{j-1}} + X_{t_{j-1}} \delta]^2 - \left( \vartheta + X_{t_{j-1}}^2 \right) \delta}{2 \left( \vartheta + X_{t_{j-1}}^2 \right)^2 \sqrt{\delta}}, \quad \tau \leq t_k \leq T.$$

The main result concerning the Pearson diffusion is the following theorem.

**Theorem 2.5.** *The one-step MLE-process  $\vartheta_{t_k, n}^*$  is consistent: for any  $\nu > 0$*

$$\mathbf{P}_{\vartheta_0} \left( \max_{N \leq k \leq n} |\vartheta_{t_k, n}^* - \vartheta_0| > \nu \right) \rightarrow 0 \quad (2.20)$$

and for all  $t \in (\tau, T]$  the convergence

$$\delta^{-1/2} (\vartheta_{t_k, n}^* - \vartheta_0) \implies \zeta_t(\vartheta_0) \quad (2.21)$$

holds. Moreover, this estimator is asymptotically efficient in the sense (2.19).

*Proof.* The consistency of the estimator is proved following the same steps as it was done in [23] and we follow the main steps of the proof of the similar result in [23], where can be found the details. We have the presentation

$$\begin{aligned} \delta^{-1/2} (\vartheta_{t_k, n}^* - \vartheta_0) &= \delta^{-1/2} (\bar{\vartheta}_N - \vartheta_0) + \frac{\Delta_{t_k, n}(\bar{\vartheta}_N, X^k)}{\mathbb{I}_{t_k, n}(\bar{\vartheta}_N)} \\ &= \delta^{-1/2} (\bar{\vartheta}_N - \vartheta_0) + \frac{\Delta_{t_k, n}(\bar{\vartheta}_N, X^k) - \Delta_{t_k, n}(\vartheta_0, X^k)}{\mathbb{I}_{t_k, n}(\vartheta_0)} \\ &\quad + \Delta_{t_k, n}(\bar{\vartheta}_N, X^k) \left( \frac{1}{\mathbb{I}_{t_k, n}(\bar{\vartheta}_N)} - \frac{1}{\mathbb{I}_{t_k, n}(\vartheta_0)} \right) + \frac{\Delta_{t_k, n}(\vartheta_0, X^k)}{\mathbb{I}_{t_k, n}(\vartheta_0)}. \end{aligned}$$

We have the stable convergence

$$\frac{\Delta_{t_k, n}(\vartheta_0, X^k)}{\mathbb{I}_{t_k, n}(\vartheta_0)} \implies \zeta_t(\vartheta_0) = \mathbb{I}_t(\vartheta_0)^{-1} \int_0^t \frac{dw(s)}{\sqrt{2}(\vartheta_0 + X_s^2)}. \quad (2.22)$$

From the continuity of Fisher information  $\mathbb{I}_{t_k, n}(\vartheta)$  and consistency of  $\bar{\vartheta}_N$  we obtain

$$\frac{1}{\mathbb{I}_{t_k, n}(\bar{\vartheta}_N)} - \frac{1}{\mathbb{I}_{t_k, n}(\vartheta_0)} \longrightarrow 0.$$

Further

$$\begin{aligned} \frac{\Delta_{t_k, n}(\bar{\vartheta}_N, X^k) - \Delta_{t_k, n}(\vartheta_0, X^k)}{\mathbb{I}_{t_k, n}(\vartheta_0)} &= \frac{\dot{\Delta}_{t_k, n}(\tilde{\vartheta}_N, X^k)(\bar{\vartheta}_N - \vartheta_0)}{\mathbb{I}_{t_k, n}(\vartheta_0)} \\ &= -\frac{\mathbb{I}_{t_k, n}(\tilde{\vartheta}_N)}{\mathbb{I}_{t_k, n}(\vartheta_0)} \delta^{-1/2} (\bar{\vartheta}_N - \vartheta_0) + o(\delta^{1/2}) = -\delta^{-1/2} (\bar{\vartheta}_N - \vartheta_0) + o(\delta^{1/2}). \end{aligned}$$

Hence (stable convergence)

$$\delta^{-1/2} (\vartheta_{t_k, n}^* - \vartheta_0) = \frac{\Delta_{t_k, n}(\vartheta_0, X^k)}{\mathbb{I}_{t_k, n}(\vartheta_0)} + o(1) \implies \zeta_t(\vartheta_0).$$

It can be shown that the moments converge too: for any  $p > 0$

$$\mathbf{E}_{\vartheta_0} |\delta^{-1/2} (\vartheta_{t_k, n}^* - \vartheta_0)|^p \longrightarrow \mathbf{E}_{\vartheta_0} |\zeta_t(\vartheta_0)|^p.$$

Moreover this convergence is uniform on  $\vartheta$ . Therefore the one-step MLE-process is asymptotically efficient estimator for polynomial loss functions.  $\blacksquare$

**Remark 2.5.** *The asymptotically efficient estimator process is constructed for values  $t \in (\tau, T]$ . Note that it is possible to have such process for all  $t \in (0, T]$ . To do this we have to consider the preliminary estimator  $\bar{\vartheta}_N$  on the interval  $[0, \tau_n]$  with  $\tau_n \rightarrow 0$  but sufficiently slowly. As it follows from the proof of the similar result in [23] we have to take  $N = n^\kappa$  with  $\kappa \in (\frac{1}{2}, 1)$ . Then  $\xi_{\tau_n} \Longrightarrow \xi_0 = \sqrt{2}(\vartheta_0 + X_0^2)\eta$ , where  $\eta \sim \mathcal{N}(0, 1)$ . Now for all  $t \in (0, T]$  we have*

$$\eta_{t,T}(\vartheta_0) = \delta^{-1/2}(\vartheta_{t_k,n}^* - \vartheta_0) \Longrightarrow \zeta_t(\vartheta_0).$$

*More detailed analysis shows that the random process  $\eta_{t,T}(\vartheta_0)$ ,  $\tau_* \leq t \leq T$  with any  $\tau_* \in (0, T)$  weakly convergences to  $\zeta_t(\vartheta_0)$ ,  $\tau_* \leq t \leq T$  in the space of continuous on  $[\tau_*, T]$  functions (see [23] for details).  $\diamond$*

# Conclusion

In this work we considered two models. In the first chapter we considered non-parametric estimation problem for the mean function of an inhomogeneous Poisson process. We obtained the following results:

**I.** Consider a kernel type estimator with the kernel  $K_n : \mathbf{R} \rightarrow \mathbf{R}$

$$\tilde{\Lambda}_n(t) = \int_0^\tau K_n(s-t)(\hat{\Lambda}_n(s) - \Lambda_*(s))ds + \Lambda_*(t).$$

Then, their Fourier coefficients with respect to the trigonometric basis will be

$$\begin{aligned}\tilde{\Lambda}_{1,n} &= \hat{\Lambda}_{1,n}, \quad \tilde{\Lambda}_{2l,n} = \sqrt{\frac{\tau}{2}}K_{2l,n}(\hat{\Lambda}_{2l,n} - \Lambda_{2l}^*) + \Lambda_{2l}^*, \\ \tilde{\Lambda}_{2l+1,n} &= \sqrt{\frac{\tau}{2}}K_{2l,n}(\hat{\Lambda}_{2l+1,n} - \Lambda_{2l+1}^*) + \Lambda_{2l+1}^*, \quad l \in \mathcal{N},\end{aligned}$$

with

$$\Lambda_l = \int_0^\tau \Lambda(t)\phi_l(t)dt, \quad \Lambda_l^* = \int_0^\tau \Lambda_*(t)\phi_l(t)dt.$$

1. For the mentioned kernel-type estimator with a kernel satisfying (1.4) (see in the text) and the condition

$$n \sup_{l \geq 1} \left| \sqrt{\frac{\tau}{2}}K_{2l,n} - 1 \right|^2 \rightarrow 0, \quad n \rightarrow +\infty,$$

then, over a  $\Lambda_*$  centered  $\mathcal{B}(R) = \{\Lambda : \int_0^\tau (\Lambda(t) - \Lambda_*(t))^2 dt \leq R\}$  ball of mean functions in  $L_2$ , the following equality holds

$$\lim_{n \rightarrow +\infty} \sup_{\Lambda \in \mathcal{B}(R)} \left( \mathbf{E}_\Lambda \|\sqrt{n}(\tilde{\Lambda}_n - \Lambda)\|^2 - \int_0^\tau \Lambda(t)dt \right) = 0.$$

2. For the kernel-type estimator with a kernel satisfying (1.4) and the condition

$$n \sup_{l \geq 1} \left| \frac{\sqrt{\frac{\tau}{2}} K_{2l,n} - 1}{\frac{2\pi l}{\tau}} \right|^2 \rightarrow 0,$$

as  $n \rightarrow +\infty$ , the following equality holds

$$\lim_{n \rightarrow +\infty} \sup_{\Lambda \in \Sigma(R)} \left( \mathbf{E}_\Lambda \| \sqrt{n}(\tilde{\Lambda}_n - \Lambda) \|^2 - \int_0^\tau \Lambda(t) dt \right) = 0,$$

where the set  $\Sigma(R)$  is defined in the Theorem 1.1.

3. Suppose that  $K(u)$ ,  $u \in [-\frac{\tau}{2}, \frac{\tau}{2}]$  is a non-negative, integrable, normalized function

$$K(u) \geq 0, u \in \left[-\frac{\tau}{2}, \frac{\tau}{2}\right], \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} K(u) du = 1,$$

which is symmetric and we continue it  $\tau$  periodically on the whole real line

$$K(u) = K(-u), \quad K(u) = K(u + \tau), \quad u \in \left[-\frac{\tau}{2}, \frac{\tau}{2}\right].$$

Let the positive sequence  $h_n \leq 1$  be such that  $h_n^2 n \rightarrow 0$ ,  $n \rightarrow +\infty$ . Then, the kernels

$$K_n(u) = \frac{1}{h_n} K\left(\frac{u}{h_n}\right) \mathbb{I}\left\{|u| \leq \frac{\tau}{2} h_n\right\}$$

satisfy (1.4) and the condition of the previous proposition.

4. For all estimators  $\bar{\Lambda}_n(t)$  of the mean function  $\Lambda(t)$ , following lower bound holds

$$\underline{\lim}_{n \rightarrow +\infty} \sup_{\Lambda \in \mathcal{F}_m^{per}(R, S)} n^{\frac{2m}{2m-1}} \left( \mathbf{E}_\Lambda \| \bar{\Lambda}_n - \Lambda \|^2 - \mathbf{E}_\Lambda \| \hat{\Lambda}_n - \Lambda \|^2 \right) \geq -\Pi.$$

5. We propose an estimator

$$\Lambda_n^*(t) = \hat{\Lambda}_{0,n} \phi_0(t) + \sum_{l=1}^{+\infty} K_{l,n} \hat{\Lambda}_{l,n} \phi_l(t),$$

where  $\{\phi_l\}_{l=1}^{+\infty}$  is the trigonometric cosine basis on  $L_2[0, \tau]$  and

$$\tilde{K}_{l,n} = \left( 1 - \left| \frac{\pi l}{\tau} \right|^m \alpha_n^* \right)_+, \quad \alpha_n^* = \left[ \frac{S \tau}{nR \pi} \frac{m}{(2m-1)(m-1)} \right]^{\frac{m}{2m-1}},$$

$$N_n = \frac{\tau}{\pi} (\alpha_n^*)^{-\frac{1}{m}} \approx C n^{\frac{1}{2m-1}}, \quad x_+ = \max(x, 0), \quad x \in \mathbf{R}.$$

Then, the estimator  $\Lambda_n^*(t)$  attains the lower bound described above, that is,

$$\lim_{n \rightarrow +\infty} \sup_{\Lambda \in \mathcal{F}_n^{per}(R,S)} n^{\frac{2m}{2m-1}} \left( \mathbf{E}_\Lambda \|\bar{\Lambda}_n - \Lambda\|^2 - \mathbf{E}_\Lambda \|\hat{\Lambda}_n - \Lambda\|^2 \right) = -\Pi.$$

That is, our proposed estimator is asymptotically second order efficient.

Results are presented in [6].

**II.** In the second chapter we considered the problem of the estimation of the solution of a FBSDE and obtained the following results: For the solution of a SDE

$$dX_t = S(t, X_t)dt + \sigma(\vartheta, t, X_t)dW(t)$$

observed at discrete moments of time we introduce the pseudo score-function (here  $A_{j-1}(\vartheta) = \sigma^2(\vartheta, t_{j-1}, X_{t_{j-1}})$ )

$$\begin{aligned} \Delta_{k,n}(\vartheta, X^k) &= \sum_{j=1}^k \dot{\ell}(\vartheta, X_{t_{j-1}}, X_{t_j}) \\ &= \frac{\left[ (X_{t_j} - X_{t_{j-1}} - S_{j-1} \delta)^2 - A_{j-1}(\theta) \delta \right] \dot{A}_{j-1}(\vartheta)}{2A_{j-1}^2(\vartheta) \sqrt{\delta}}, \end{aligned}$$

and the random Fisher information

$$\mathbb{I}_\tau(\vartheta_0) = 2 \int_0^\tau \frac{\dot{\sigma}(\vartheta_0, s, X_s) \dot{\sigma}(\vartheta_0, s, X_s)^\mathbb{T}}{\sigma^2(\vartheta_0, s, X_s)} ds.$$

For any  $t \in [\tau, T]$  define  $k$  by the condition  $t_k \leq t < t_{k+1}$  and the one-step PMLE-process by the relation

$$\vartheta_{k,n}^* = \hat{\vartheta}_{\tau,n} + \sqrt{\delta} \mathbb{I}_{k,n}(\hat{\vartheta}_{\tau,n})^{-1} \Delta_{k,n}(\hat{\vartheta}_{\tau,n}, X^k), \quad k = N+1, \dots, n.$$

1. The one-step MLE-process  $\vartheta_{k,n}^*, k = N+1, \dots, n$  is consistent and asymptotically conditionally normal (stable convergence)

$$\delta^{-1/2} (\vartheta_{k,n}^* - \vartheta_0) \Longrightarrow \xi_t(\vartheta_0)$$

and is asymptotically efficient for  $t \in [\tau_*, T]$  where  $\tau < \tau_* < T$  and a bounded loss functions.



2. Suppose that the coefficients of the diffusion process satisfies  $\mathcal{R}$  conditions, then, for the loss function  $\ell(u) = |u|^p$ ,  $p > 0$ , the following lower bound is true

$$\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow +\infty} \sup_{|\vartheta - \vartheta_0| < \varepsilon} \mathbf{E}_{\vartheta} \ell(\delta^{-1/2} (\bar{Y}_{t_k, n} - Y_{t_k})) \geq \mathbf{E}_{\vartheta_0} \ell(\dot{u}(\vartheta_0, t, X_t) \xi_t(\vartheta_0)).$$

Here  $u(\vartheta, t, x)$  satisfies the equation

$$\frac{\partial u}{\partial t} + S(t, x) \frac{\partial u}{\partial x} + \frac{\sigma(\vartheta, t, x)^2}{2} \frac{\partial^2 u}{\partial x^2} = -f\left(t, x, u, \sigma(\vartheta, t, x) \frac{\partial u}{\partial x}\right).$$

3. Suppose that the conditions of regularity hold, then the estimators

$Y_{t_k, n}^* = u(t_k, X_{t_k}, \vartheta_{k, n}^*)$ ,  $Z_{t_k, n}^* = u'_x(t_k, X_{t_k}, \vartheta_{k, n}^*) \sigma(t_k, X_{t_k}, \vartheta_{k, n}^*)$ ,  $t_k \in [\tau, T]$ , are consistent  $Y_{t_k, n}^* \rightarrow Y_{t_k}$ ,  $Z_{t_k, n}^* \rightarrow Z_{t_k}$ , and are asymptotically conditionally normal (stable convergence)

$$\begin{aligned} \delta^{-1/2} (Y_{t_k, n}^* - Y_{t_k}) &\Longrightarrow \langle \dot{u}(t, X_t, \vartheta_0), \xi_t(\vartheta_0) \rangle, \\ \delta^{-1/2} (Z_{t_k, n}^* - Z_{t_k}) &\Longrightarrow \sigma(t, X_t, \vartheta_0) \langle \dot{u}'_x(t, X_t, \vartheta_0), \xi_t(\vartheta_0) \rangle \\ &\quad + u'_x(t, X_t, \vartheta_0) \langle \dot{\sigma}(t, X_t, \vartheta_0), \xi_t(\vartheta_0) \rangle. \end{aligned}$$

4. Consider a Pearson diffusion

$$dX_t = -X_t dt + \sqrt{\vartheta + X_t^2} dW_t, \quad X_0, \quad 0 \leq t \leq T.$$

For this model, using the preliminary estimator

$$\bar{\vartheta}_N = \frac{n}{TN} \left[ X_{t_N}^2 - X_0^2 - 2 \sum_{j=1}^N X_{t_{j-1}} [X_{t_j} - X_{t_{j-1}}] - \sum_{j=1}^N X_{t_{j-1}}^2 \delta \right],$$

we propose the one-step MLE process

$$\vartheta_{t_k, n}^* = \bar{\vartheta}_N + \sqrt{\delta} \sum_{j=1}^k \frac{[X_{t_j} - X_{t_{j-1}} + X_{t_{j-1}} \delta]^2 - (\bar{\vartheta}_N + X_{t_{j-1}}^2) \delta}{2 \mathbb{I}_{t_k, n}(\bar{\vartheta}_N) (\bar{\vartheta}_N + X_{t_{j-1}}^2)^2 \sqrt{\delta}}, \quad \tau \leq t_k \leq T.$$

and prove that the one-step MLE-process  $\vartheta_{t_k, n}^*$  is consistent: for any  $\nu > 0$

$$\mathbf{P}_{\vartheta_0} \left( \max_{N \leq k \leq n} |\vartheta_{t_k, n}^* - \vartheta_0| > \nu \right) \rightarrow 0$$

and for all  $t \in (\tau, T]$  the convergence

$$\delta^{-1/2} (\vartheta_{t_k, n}^* - \vartheta_0) \Longrightarrow \zeta_t(\vartheta_0)$$

holds. Moreover, this estimator is asymptotically efficient.

The results are presented in [7],[8].

# Appendix

*Proof.* [Theorem 1.2] We are going to use an important equality, that  $\mu$ -almost everywhere ( $\mu$  is the Lebesgue measure on  $\Theta$  and  $\mu(\Theta) < +\infty$ )

$$\mathbf{E}_\theta \left( \frac{\frac{\partial f(X_j, \theta)}{\partial \theta_k}}{f(X_j, \theta)} \right) = 0, \quad j \in \{1, \dots, n\}, k \in \mathcal{N}.$$

To prove this, first note that

$$\int_{\mathcal{X}} \int_{\Theta} \left| \frac{\partial f(X, \theta)}{\partial \theta_k} \right| d\theta \nu(dx) < +\infty.$$

Indeed, for  $\nu$ -almost all  $x \in \mathcal{X}$  the function  $\theta_k \mapsto f(x, \theta)$  is absolutely continuous on  $[\alpha_k, \beta_k]$ , hence the derivative  $\frac{\partial f(x, \theta)}{\partial \theta_k}$  exists almost everywhere on  $[\alpha_k, \beta_k]$ . Choose  $x \in \mathcal{X}$  so that  $\theta_k \mapsto f(x, \theta)$  is absolutely continuous, then

$$\{\theta_k \in [\alpha_k, \beta_k] : f(x, \theta) = 0\} \subset \{\theta_k \in [\alpha_k, \beta_k] : \frac{\partial f(x, \theta)}{\partial \theta_k} = 0\} \cup A_x,$$

where  $A_x \subset [\alpha_k, \beta_k]$  is of Lebesgue measure 0. Then, denoting,  $B = \{(x, \theta) : f(x, \theta) \neq 0\}$  and using the Tonelli theorem to write the double integral as an iterated integral

$$\begin{aligned} \int_{\mathcal{X}} \int_{\Theta} \left| \frac{\partial f(x, \theta)}{\partial \theta_k} \right| d\theta \nu(dx) &= \int_{\mathcal{X}} \int_{\Theta} \left| \frac{\partial f(x, \theta)}{\partial \theta_k} \right| \mathbb{1}(B) d\theta \nu(dx) = \\ \int_{\Theta} \left( \int_{\mathcal{X}} \left| \frac{\partial f(x, \theta)}{\partial \theta_k} \right| \mathbb{1}(B) \mathbf{P}_\theta(dx) \right) d\theta &= \int_{\Theta} \int_{\mathcal{X}} \left| \frac{\partial f(x, \theta)}{\partial \theta_k} \right| \mathbb{1}(B) \mathbf{P}_\theta(dx) d\theta. \end{aligned}$$

Applying the Cauchy-Schwarz inequality to the last integral gives us

$$\int_{\mathcal{X}} \int_{\Theta} \left| \frac{\partial f(x, \theta)}{\partial \theta_k} \right| d\theta \nu(dx) \leq (\mathbf{E}_{\mathbf{Q}} \mathbf{I}_k(\theta))^{\frac{1}{2}} \sqrt{\mu(\Theta)} < +\infty.$$

Therefore, we can use the Fubini theorem, according to which for all  $a, b \in [\alpha_k, \beta_k]$

$$\begin{aligned} \int_a^b \int_{\mathcal{X}} \frac{\partial f(\theta, x)}{\partial \theta_k} \nu(dx) d\theta_k &= \int_{\mathcal{X}} \left( \int_a^b \frac{\partial f(\theta, x)}{\partial \theta_k} d\theta_k \right) \nu(dx) = \\ &= \int_{\mathcal{X}} f(\theta, x)|_{\theta_k=b} \nu(dx) - \int_{\mathcal{X}} f(\theta, x)|_{\theta_k=a} \nu(dx) = 0, \end{aligned}$$

since  $f(\theta, x)$  is a density for all  $\theta \in \Theta$ . Here  $a, b$  and  $\theta_k$  were arbitrary, therefore for  $\mu$ -almost all  $\theta \in \Theta$

$$\mathbf{E}_{\theta} \left( \frac{\frac{\partial f(X_j, \theta)}{\partial \theta_k}}{f(X_j, \theta)} \right) = \int_{\mathcal{X}} \frac{\partial f(\theta, x)}{\partial \theta_k} \nu(dx) = 0.$$

The function  $p_k(\theta) \prod_{j=1}^n f(X_j, \theta)$  is absolutely continuous in  $\theta_k$  and since  $p_k(\alpha_k) = p_k(\beta_k) = 0$ , hence

$$\int_{\alpha_k}^{\beta_k} \frac{\partial}{\partial \theta_k} \left( p_k(\theta) \prod_{j=1}^n f(X_j, \theta) \right) d\theta_k = 0,$$

by same arguments, using the integration by parts formula,

$$\int_{\alpha_k}^{\beta_k} \psi(\theta) \frac{\partial}{\partial \theta_k} \left( p_k(\theta) \prod_{j=1}^n f(X_j, \theta) \right) d\theta_k = - \int_{\alpha_k}^{\beta_k} \frac{\partial \psi(\theta)}{\partial \theta_k} p_k(\theta) \prod_{j=1}^n f(X_j, \theta) d\theta_k.$$

Combining these two equalities, for the estimator  $\psi_n = \psi_n(\mathbf{X}^n)$  the equality holds

$$\begin{aligned} \int_{\alpha_k}^{\beta_k} (\psi_n - \psi(\theta)) \frac{\partial}{\partial \theta_k} \left( p_k(\theta_k) \prod_{j=1}^n f(X_j, \theta) \right) d\theta_k = \\ \int_{\alpha_k}^{\beta_k} \frac{\partial \psi(\theta)}{\partial \theta_k} p_k(\theta_k) \prod_{j=1}^n f(X_j, \theta) d\theta_k. \end{aligned}$$

Consider  $\theta^{(k)} = (\theta_1, \dots, \theta_{k-1}, \theta_{k+1}, \dots)$ , the parameter  $\theta$  without the  $\theta_k$  coordinate, the corresponding set  $\Theta^{(k)} = \{\theta^{(k)}, \theta_i \in [\alpha_i, \beta_i], i \neq k\}$ , and its distribution  $\mathbf{Q}^{(k)}(d\theta^{(k)})$ . Integrating the last equality w.r.t. the measure  $\mathbf{Q}^{(k)}(d\theta^{(k)}) \nu^n(dx)$  gives us

$$\begin{aligned} \int_{\mathcal{X}^n} \int_{\Theta^{(k)}} \int_{\alpha_k}^{\beta_k} (\psi_n - \psi(\theta)) \frac{\partial}{\partial \theta_k} \left( p_k(\theta_k) \prod_{j=1}^n f(X_j, \theta) \right) d\theta_k \mathbf{Q}^{(k)}(d\theta^{(k)}) \nu^n(dx) = \\ \int_{\mathcal{X}^n} \int_{\Theta^{(k)}} \int_{\alpha_k}^{\beta_k} \frac{\partial \psi(\theta)}{\partial \theta_k} p_k(\theta_k) \prod_{j=1}^n f(X_j, \theta) d\theta_k \mathbf{Q}^{(k)}(d\theta^{(k)}) \nu^n(dx). \end{aligned}$$

Using the Foubini theorem, we can show that the right-hand side equals to

$$\int_{\mathcal{X}^n} \int_{\Theta^{(k)}} \int_{\alpha_k}^{\beta_k} \frac{\partial \psi(\theta)}{\partial \theta_k} p_k(\theta_k) \prod_{j=1}^n f(X_j, \theta) d\theta_k \mathbf{Q}^{(k)}(d\theta^{(k)}) \nu^n(dx) = \mathbf{E}_{\mathbf{Q}} \frac{\partial \psi(\theta)}{\partial \theta_k}.$$

Modifying also the left-hand side, we finally can write

$$\int_{\mathcal{X}^n} \int_{\Theta} (\psi_n - \psi(\theta)) \frac{\frac{\partial}{\partial \theta_k} \left( p_k(\theta_k) \prod_{j=1}^n f(X_j, \theta) \right)}{p_k(\theta_k) \prod_{j=1}^n f(X_j, \theta)} \mathbf{Q}(d\theta) \mathbf{P}_{\theta}^n(dx) = \mathbf{E}_{\mathbf{Q}} \frac{\partial \psi(\theta)}{\partial \theta_k}.$$

Applying the Cauchy-Schwarz inequality we deduce

$$\begin{aligned} \int_{\mathcal{X}^n} \int_{\Theta} (\psi_n - \psi(\theta))^2 \mathbf{Q}(d\theta) \mathbf{P}_{\theta}^n(dx) \int_{\mathcal{X}^n} \int_{\Theta} Y^2 \mathbf{Q}(d\theta) \mathbf{P}_{\theta}^n(dx) &\geq \\ &\geq \left[ \mathbf{E}_{\mathbf{Q}} \frac{\partial \psi(\theta)}{\partial \theta_k} \right]^2, \end{aligned}$$

with the notation

$$Y = \frac{\frac{\partial}{\partial \theta_k} \left( p_k(\theta_k) \prod_{j=1}^n f(X_j, \theta) \right)}{p_k(\theta_k) \prod_{j=1}^n f(X_j, \theta)}.$$

First, calculate the following derivatives

$$\begin{aligned} &\frac{\partial}{\partial \theta_k} \left( p_k(\theta_k) \prod_{j=1}^n f(X_j, \theta) \right) = \\ &= p'_k(\theta_k) \prod_{j=1}^n f(X_j, \theta) + p_k(\theta_k) \sum_{j=1}^n f(X_1, \theta) \cdots \frac{\partial f(X_j, \theta)}{\partial \theta_k} \cdots f(X_n, \theta) = \\ &= p_k(\theta_k) \prod_{j=1}^n f(X_j, \theta) \left( \frac{p'_k(\theta_k)}{p_k(\theta_k)} + \sum_{j=1}^n \frac{\frac{\partial f(X_j, \theta)}{\partial \theta_k}}{f(X_j, \theta)} \right). \end{aligned}$$

Further, the following mathematical expectation equals

$$\begin{aligned} &\mathbb{E} \left( \frac{p'_k(\theta_k)}{p_k(\theta_k)} + \sum_{j=1}^n \frac{\frac{\partial f(X_j, \theta)}{\partial \theta_k}}{f(X_j, \theta)} \right)^2 = \\ &\mathbb{E} \left( \frac{p'_k(\theta)}{p_k(\theta)} \right)^2 + \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} \frac{\frac{\partial f(X_i, \theta)}{\partial \theta_k}}{f(X_i, \theta)} \frac{\frac{\partial f(X_j, \theta)}{\partial \theta_k}}{f(X_j, \theta)} + \sum_{j=1}^n \mathbb{E} \frac{p'_k(\theta)}{p_k(\theta)} \frac{\frac{\partial f(X_j, \theta)}{\partial \theta_k}}{f(X_j, \theta)}. \end{aligned}$$

Here, we use the following identities

$$\begin{aligned} \mathbf{E}_\theta \left( \frac{\frac{\partial f(X_i, \theta)}{\partial \theta_k}}{f(X_i, \theta)} \frac{\frac{\partial f(X_j, \theta)}{\partial \theta_k}}{f(X_j, \theta)} \right) &= \mathbf{E}_\theta \left( \frac{\frac{\partial f(X_i, \theta)}{\partial \theta_k}}{f(X_i, \theta)} \right) \mathbf{E}_\theta \left( \frac{\frac{\partial f(X_j, \theta)}{\partial \theta_k}}{f(X_j, \theta)} \right) = 0, \\ \mathbf{E}_\theta \left( \frac{p'_k(\theta)}{p_k(\theta)} \frac{\frac{\partial f(X_j, \theta)}{\partial \theta_k}}{f(X_j, \theta)} \right) &= \int_{\Theta} \frac{p'_k(\theta)}{p_k(\theta)} \mathbf{E}_\theta \left( \frac{\frac{\partial f(X_j, \theta)}{\partial \theta_k}}{f(X_j, \theta)} \right) \mathbf{Q}(d\theta) = 0, \\ \mathbb{E} \left( \frac{\frac{\partial f(X_j, \theta)}{\partial \theta_k}}{f(X_j, \theta)} \right)^2 &= \mathbf{E}_{\mathbf{Q}} \mathbf{I}_k(\theta), \quad \mathbb{E} \left( \frac{p'_k(\theta)}{p_k(\theta)} \right)^2 = \mathbf{E}_{\mathbf{Q}} \left( \frac{p'_k(\theta)}{p_k(\theta)} \right)^2 = \mathbf{I}_k. \end{aligned}$$

These identities are true because of the basic identity  $\mathbf{E}_\theta \left( \frac{\frac{\partial f(X_i, \theta)}{\partial \theta_k}}{f(X_i, \theta)} \right) = 0$ . Combining all these facts we get

$$\mathbb{E} \left[ \frac{\frac{\partial}{\partial \theta_k} \left( p_k(\theta_k) \prod_{j=1}^n f(X_j, \theta) \right)}{p_k(\theta_k) \prod_{j=1}^n f(X_j, \theta)} \right]^2 = n \mathbf{E}_{\mathbf{Q}} \mathbf{I}_k(\theta) + \mathbf{I}_k.$$

Finally, in the Cauchy-Schwarz inequality above, we can replace

$$\int_{\mathcal{X}^n} \int_{\Theta} Y^2 \mathbf{Q}(d\theta) \mathbf{P}_\theta^n(dx) = \mathbb{E} Y^2 = n \mathbf{E}_{\mathbf{Q}} \mathbf{I}_k(\theta) + \mathbf{I}_k,$$

which completes the proof. ■

*Proof.* [Theorem 1.1]. We want to proof that for all estimators  $\bar{\Lambda}_n$

$$\liminf_{n \rightarrow +\infty} \sup_{\Lambda \in \mathcal{F}_m^{per}(R, S)} \left( \mathbf{E}_\Lambda \| \sqrt{n}(\bar{\Lambda}_n - \Lambda) \|^2 - \int_0^\tau \Lambda(t) dt \right) \geq 0,$$

where

$$\mathcal{F}_m^{per}(R, S) = \left\{ \Lambda(\cdot) : \int_0^\tau [\Lambda^{(m)}(t)]^2 dt \leq R, \Lambda(0) = 0, \Lambda(\tau) = S \right\}, m \in \mathcal{Z}_+.$$

Periodicity means that the intensity function  $\lambda(\cdot)$  is periodic (hence the equality of its values and the values of its derivatives on the endpoints of the interval  $[0, \tau]$ ). For  $m = 0$  we get an  $L_2$  ball

$$\mathcal{F}_0^{per}(R, S) = \left\{ \Lambda(\cdot) : \int_0^\tau \Lambda^2(t) dt \leq R, \Lambda(0) = 0, \Lambda(\tau) = S \right\}.$$

For  $m > 1$  the statement of this theorem follows from the Proposition 1.3. From Lemma 1.1 we know that  $\Lambda \in \mathcal{F}_m^{per}(R, S)$  if and only if its Fourier coefficients w.r.t. the cosine basis satisfy

$$\sum_{l=1}^{+\infty} \left(\frac{\pi l}{\tau}\right)^{2m} \Lambda_l^2 \leq R, \quad \Lambda(\tau) = S.$$

Using the inequalities

$$\sum_{l=1}^{+\infty} \Lambda_l^2 = \sum_{l=1}^{+\infty} \left(\frac{\tau}{\pi l}\right)^2 \left(\frac{\pi l}{\tau}\right)^2 \Lambda_l^2 \leq \left(\frac{\tau}{\pi}\right)^2 \sum_{l=1}^{+\infty} \left(\frac{\pi l}{\tau}\right)^2 \Lambda_l^2 \leq \left(\frac{\tau}{\pi}\right)^4 \sum_{l=1}^{+\infty} \left(\frac{\pi l}{\tau}\right)^4 \Lambda_l^2$$

therefore  $\mathcal{F}_2^{per}(R, S) \subset \mathcal{F}_1^{per}(R_1, S) \subset \mathcal{F}_0^{per}(R_2, S)$ , which ensures that for the functional

$$F_n(\bar{\Lambda}_n, \Lambda) = \mathbf{E}_\Lambda \|\sqrt{n}(\bar{\Lambda}_n - \Lambda)\|^2 - \int_0^\tau \Lambda(t) dt$$

the following is true

$$\sup_{\Lambda \in \mathcal{F}_0^{per}(R_2, S)} F_n(\bar{\Lambda}_n, \Lambda) \geq \sup_{\Lambda \in \mathcal{F}_1^{per}(R_1, S)} F_n(\bar{\Lambda}_n, \Lambda) \geq \sup_{\Lambda \in \mathcal{F}_2^{per}(R, S)} F_n(\bar{\Lambda}_n, \Lambda),$$

and since the Theorem is true for  $m > 1$  then, the Theorem is true for  $m = 0$  and  $m = 1$ . ■



# Bibliography

- [1] Belister E.N. and Levit B.Ya., (1995) On minimax filtering over ellipsoids. *Math. Methods of Statist.*, 4, 259-273.
- [2] Bismut J.M., (1973) Conjugate convex functions in optimal stochastic control. *J. Math. Anal. Appl.*, 44, 384-404.
- [3] Dalalyan A.S. and Kutoyants Yu.A., (2004) On second order minimax estimation of invariant density for ergodic diffusion. *Statistics & Decisions*, 22(1), 17-42.
- [4] Dohnal G., (1987) On estimating the diffusion coefficient. *J. Appl. Probab.*, 24(1), 105-114.
- [5] El Karoui N., Peng S. and Quenez M., (1997) Backward stochastic differential equations in finance. *Math. Fin.*, 7, 1-71.
- [6] Gasparyan S.B., (2015) Second order asymptotical efficiency for a Poisson process. *J. Contemp. Math. Anal.*, 50(2), 98-106.
- [7] Gasparyan S.B. and Kutoyants Yu.A., (2015) On approximation of the BSDE with unknown volatility in forward equation. *Armenian J. Mathem.*, 7(1), 59-79.
- [8] Gasparyan S.B. and Kutoyants Yu.A., (2015) An example of one-step MLE-process in volatility estimation problem. *Izvestia of NAS, Seria Math.*, 50(3), 71-76.
- [9] Genon-Catalot V. and Jacod J., (1993) On the estimation of the diffusion coefficient for multi-dimensional diffusion. *Ann. IHP, Sec. B*, 29(1), 119-151.
- [10] Gill R.D. and Levit B.Ya., (1995) Applications of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, 1(1-2), 59-79.
- [11] Gobet E., (2001) Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach. *Bernoulli*, 7(6), 889-912.



- 
- [12] Golubev G.K. and Härdle W., (2000) Second order minimax estimation in partial linear models. *Math. Methods Statist.*, 9(2), 160-175.
- [13] Golubev G.K. and Levit B.Ya., (1996) On the second order minimax estimation of distribution functions. *Math. Methods Statist.*, 5(1), 1-31.
- [14] Ibragimov I.A. and Khasminskii R.Z., (1981) *Statistical Estimation. Asymptotic Theory*. Springer, N. Y.
- [15] Jeganathan P., (1982) On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhya A*, 173-212.
- [16] Jeganathan P., (1983) Some asymptotic properties of risk functions when the limit of the experiment is mixed normal. *Sankhya A, Pt.1*, 45, 66-87.
- [17] Johnstone I. M., (2015) *Gaussian Estimation: Sequence and Wavelet Models*. Book draft, version - September 8.
- [18] Kingman J.F.C., (1992) *Poisson Processes*. Oxford University press.
- [19] Kamatani K. and Uchida M., (2015) Hybrid multi-step estimators for stochastic differential equations based on sampled data. *Stat. Inference Stoch. Process.*, 18(2), 177-204.
- [20] Kutoyants Yu. A., (2016) *Introduction to Statistics of Poisson Processes*. To appear.
- [21] Kutoyants Yu.A., (1998) *Statistical Inference for Spatial Poisson Processes*. Lecture Notes in Statistics 134, Springer, N. Y.
- [22] Kutoyants Yu.A., (2014) Approximation of the backward stochastic differential equation. Small noise, large samples and high frequency cases. *Proceedings of the Steklov Institute of Mathematics*, 287, 133-154.
- [23] Kutoyants Yu.A., (2015) On multi-step MLE-processes for ergodic diffusion. *Submitted*.
- [24] Kutoyants Yu.A. and Zhou L., (2014) On approximation of the backward stochastic differential equation. *J. Stat. Plann. Infer.*, 150, 111-123.
- [25] Le Cam L., (1956) On the asymptotic theory of estimation and testing hypotheses. *Proc. 3rd Berkeley Symposium I*, 355-368.
- [26] Lehmann E.L. and Romano J.P., (2005) *Testing Statistical Hypotheses*. (3rd ed.) Springer, N. Y.
- [27] Levanony D., Shwartz A. and Zeitouni O., (1994) Recursive identification in continuous-time stochastic process. *Stochastic Process. Appl.*, 49, 245-275.

- 
- [28] Levit B.Y., (1978) Infinite-dimensional informational bounds. *Theory Probab. Appl.*, 23(2), 388–394.
- [29] Liptser R. and Shiryaev A.N., (2005) *Statistics of Random Processes*. Springer, N. Y.
- [30] Ma J. and Yong J., (1999) *Forward-Backward Stochastic Differential Equations and their Applications*. Lecture Notes in Mathematics. Springer, Berlin.
- [31] Nussbaum M., (1999) Minimax risk: Pinsker bound. In *Encyclopedia of Statistical Sciences*, (Update volume 3), S. Kotz (Ed.), New York: Wiley, 451-460.
- [32] Oksendal B., (1992) *Stochastic Differential Equations: An Introduction with Applications*. Springer, N. Y.
- [33] Pardoux E. and Peng S., (1990) Adapted solution of a backward stochastic differential equation. *System Control Letter*, 14, 55-61.
- [34] Pardoux E. and Peng S., (1992) Backward stochastic differential equation and quasilinear parabolic differential equations. In *Stochastic Partial Differential equations and Their Applications, Lecture Notes in Control and Information Sciences*, 176, 200-217.
- [35] Pardoux E. and Răscanu S., (2014) *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations*. Springer, N. Y.
- [36] Pinsker M.S., (1980) Optimal filtering of square-integrable signals in Gaussian noise. *Probl. Inf. Transm.*, 16(2), 120-133.
- [37] Reynaud-Bouret, P. (2003) Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probab. Theory Rel.* 126(1), 103-153.
- [38] Sørensen M., (2009) Estimating functions for diffusion-type processes. In *Statistical Methods for Stochastic Differential Equations*, Kessler, Lindar and Sørensen (Ed's), Chapman & Hall/CRC, Boca Raton, 1-107.
- [39] Tikhomirov V.M., (1976) *Some Problems in Approximation Theory*. Moscow Univ. Press, Moscow. (In Russian).
- [40] Tsybakov A.B., (2009) *Introduction to Nonparametric Estimation*. Springer Series in Statistics, Springer, N. Y.
- [41] Wong E., (1971) Representation of martingales, quadratic variation and applications. *SIAM Journal on Control*, 9, 4, 621-633.

# Thèse de Doctorat

Samvel GASPARYAN

Deux problèmes d'estimation statistique pour les processus stochastiques

Two problems of statistical estimation for stochastic processes

## Résumé

Le travail est consacré aux questions de la statistique des processus stochastiques. Particulièrement, on considère deux problèmes d'estimation. Le premier chapitre se concentre sur le problème d'estimation non-paramétrique pour le processus de Poisson non-homogène. On estime la fonction moyenne de ce processus, donc le problème est dans le domaine d'estimation non-paramétrique. On commence par la définition de l'efficacité asymptotique dans les problèmes non-paramétriques et on procède à exploration de l'existence des estimateurs asymptotiquement efficaces. On prend en considération la classe des estimateurs à noyau. Dans la thèse il est démontré que sous les conditions sur les coefficients du noyau par rapport à une base trigonométrique, on a l'efficacité asymptotique dans le sens minimax sur les ensembles divers. Les résultats obtenus soulignent le phénomène qu'en imposant des conditions de régularité sur la fonction inconnue, on peut élargir la classe des estimateurs asymptotiquement efficaces. Pour comparer les estimateurs asymptotiquement efficaces (du premier ordre), on démontre une inégalité qui nous permet de trouver un estimateur qui est asymptotiquement efficace du second ordre. On calcule aussi la vitesse de convergence pour cet estimateur, qui dépend de la régularité de la fonction inconnue et finalement on calcule la valeur minimale de la variance asymptotique pour cet estimateur. Cette valeur joue le même rôle dans l'estimation du second ordre que la constante de Pinsker dans le problème d'estimation de la densité ou encore l'information de Fisher dans les problèmes d'estimation paramétrique.

Le deuxième chapitre est dédié au problème de l'estimation de la solution d'une équation différentielle stochastique rétrograde (EDSR). On observe un processus de diffusion qui est donnée par son équation différentielle stochastique dont le coefficient de la diffusion dépend d'un paramètre inconnu. Les observations sont discrètes. Pour estimer la solution de l'EDSR on a besoin d'un estimateur-processus pour le paramètre, qui, chaque instant n'utilise que la partie des observations disponible. Dans la littérature il existe une méthode de construction, qui minimise une fonctionnelle. On ne pouvait pas utiliser cet estimateur, car le calcul serait irréalisable. Dans le travail nous avons proposé un estimateur-processus qui a la forme simple et peut être facilement calculé. Cet estimateur-processus est un estimateur asymptotiquement efficace et en utilisant cet estimateur on estime la solution de l'EDSR de manière efficace aussi.

**Mots clés** estimation non-paramétrique, efficacité asymptotique, processus stochastiques

## Abstract

This work is devoted to the questions of the statistics of stochastic processes. Particularly, the first chapter is devoted to a non-parametric estimation problem for an inhomogeneous Poisson process. The estimation problem is non-parametric due to the fact that we estimate the mean function. We start with the definition of the asymptotic efficiency in non-parametric estimation problems and continue with examination of the existence of asymptotically efficient estimators. We consider a class of kernel-type estimators. In the thesis we prove that under some conditions on the coefficients of the kernel with respect to a trigonometric basis we have asymptotic efficiency in minimax sense over various sets. The obtained results highlight the phenomenon that imposing regularity conditions on the unknown function, we can widen the class of asymptotically efficient estimators. To compare these (first order) efficient estimators, we prove an inequality which allows us to find an estimator which is asymptotically efficient of second order. We calculate also the rate of convergence of this estimator, which depends on the regularity of the unknown function, and finally the minimal value of the asymptotic variance for this estimator is calculated. This value plays the same role in the second order estimation as the Pinsker constant in the density estimation problem or the Fisher information in parametric estimation problems.

The second chapter is dedicated to a problem of estimation of the solution of a Backward Stochastic Differential Equation (BSDE). We observe a diffusion process which is given by its stochastic differential equation with the diffusion coefficient depending on an unknown parameter. The observations are discrete. To estimate the solution of a BSDE, we need an estimator-process for a parameter, which, for each given time, uses only the available part of observations. In the literature there exists a method of construction, which minimizes a functional. We could not use this estimator, because the calculations would not be feasible. We propose an estimator-process which has a simple form and can be easily computed. Using this estimator we estimate the solution of a BSDE in an asymptotically efficient way.

**Key Words** non-parametric estimation, asymptotic efficiency, stochastic processes